US006133417A

United States Patent [19]

Koch et al.

[11] **Patent Number: 6,133,417**

[45] **Date of Patent: Oct. 17, 2000**

[54] **CYTOCHROME P-450 MONOOXYGENASES**

[75] Inventors: **Birgit Maria Koch**, Vanløse; **Ole Sibbesen**, Roskilde; **Barbara Ann Halkier**, Copenhagen V; **Birger Lindberg Møller**, Brønshøj, all of Denmark

[73] Assignees: **Novartis Finance Corporation**, New York, N.Y.; **Royal Veterinary Agricultural University**, Copenhagen, Denmark

[21] Appl. No.: **09/256,797**

[22] Filed: **Feb. 24, 1999**

Related U.S. Application Data

[62] Division of application No. 08/656,177, filed as application No. PCT/EP94/03938, Nov. 28, 1994, Pat. No. 5,882,851.

[30] **Foreign Application Priority Data**

Dec. 8, 1993 [EP] European Pat. Off. 93810860

[51] **Int. Cl.**[7] **C12Q 1/00**; C12Q 1/26; G01N 33/53; C07H 19/00

[52] **U.S. Cl.** **530/300**; 530/350; 435/4; 435/440; 435/468; 800/278; 800/288

[58] **Field of Search** 530/300, 350; 435/4, 440, 468; 800/278, 288

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,882,851 3/1999 Koch et al. 435/4

FOREIGN PATENT DOCUMENTS

0 281 245 A2	2/1988	European Pat. Off. .
0 522 880 A3	7/1992	European Pat. Off. .
WO 91/10745	1/1991	WIPO .
WO 93/04174	8/1992	WIPO .
WO 93/21326	4/1993	WIPO .
WO 93/21326	10/1993	WIPO .

OTHER PUBLICATIONS

Sibbesen et al. "Isolation of the heme-thiolate enzyme cytochrome P-450Tyr, which catalysis the committed step in the biosynthesis of teh cyanogenic glucoside dhurrin in Sorghum bicolor (L.)Moench" Proc. Natl. Acad. Sci. USA, vol. 91, pp. 9740–9744, Oct. 1994.

Gabriac, et al., "Purification and Immunocharacterization of a Plant Cytochrome P450: The Cinnamic Acid 4–Hydroxylase", *Archives of Biochemistry and Biophysics,* 288(1): 302–309 (1991).

Halkier et al., "The Biosynthesis of Cyanogenic Glucosides in Higher Plants", *The Journal of Biological Chemistry,* 264(33): 19487–19494 (1989).

Halkier, et al., "Biosynthesis of the Cyanogenic Glucoside Dhurrin in Seedlings of *Sorghum bicolor* (L.) Moench and Partial Purification of the Enzyme System Involved", *Plant Physiol.,* 90: 1552–1559 (1989).

Halkier, et al., "Cyanogenic glucosides: the biosynthetic pathway and the enzyme system involved", *Department of Plant Physioology, Royal Veterinary and Agricultural University,* 40: 49–66 (1988).

Halkier, et al., "The Biosynthesis of Cyanogenic Glucosides in Higher Plants", *The Journal of Biological Chemistry,* 265(34): 21114–21121 (1990).

Halkier, et al., "Involvement of Cytochrome P–450 in the Biosynthesis of Dhurrin in *Sorghum bicolor* (L.) Moench", *Plant Physiol.,* 96: 10–17 (1991).

Halkier, et al., "2–Nitro–3–(p–hydroxyphenyl)propionate and aci–1–nitro–2–(p–hydroxyphenyl)ethane, two intermediates in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench", *Proc. Natl. Acad. Sci.,* 88: 487–491 (1991).

Koch, et al., "The biosynthesis of cyanogenic glucosides in seedlings of cassava (*Manihot esculenta* Crantez),", *Archives of Biochemistry and Biophysics,* 292: 141–150 (1992).

Sibbesen et al., "Purification of the Hydroxylating Enzyme System Involved in the Biosynthesis of the Cyanogenic Glucoside Dhurrin in *Sorghum Bicolor* (l.) Moench", *Biochemistry and Biophysics of Cytochrome P450,* 232 (1991).

Primary Examiner—Jezia Riley
Attorney, Agent, or Firm—J. Timothy Meigs

[57] **ABSTRACT**

New cytochrome P-450 dependent monooxygenases and DNA molecules encoding these monooxygenases are provided, which are able to catalyze the biosynthetic pathway from amino acids to their corresponding cyanohydrins, the presursors of the cyanogenic glycosides, or to glucosinolates. Moreover, the invention provides methods for obtaining DNA molecules according to the invention and methods for obtaining transgenic plants resistant to insects, acarids, or nematodes or plants with improved nutritive value.

37 Claims, No Drawings

CYTOCHROME P-450 MONOOXYGENASES

This is a divisional of application Ser. No. 08/656,177, filed Aug. 8, 1996, now U.S. Pat. No. 5,882,851 which is a § 371 of PCT/EP94/03938, filed Nov. 28, 1994 (published Jun. 15, 1995, as WO 95/16041), which claims priority of European Patent Application No. 93810860.2, filed Dec. 8, 1993.

The present invention relates to genetic engineering in plants using recombinant DNA technology in general and to enzymes involved in the biosynthesis of cyanogenic glycosides and glucosinolates and genes encoding these enzymes in particular. The proteins and genes according to the invention can be used to improve the nutritive value or pest resistance of plants.

Cyanogenic glycosides constitute secondary plant metabolites in more than 2000 plant species. In some instances they are the source of HCN which can render a plant toxic if it is taken as food. For example the tubers of the cyanogenic crop cassava (*Manihot esculenta*) constitute an important staple food in tropical areas. However, the cyanogenic glycosides present in the tubers may cause cyanide poisoning in humans due to insufficiently processed cassava products. Other plant species whose enzymatic production of HCN accounts for their potential toxicity if taken in excess as food or used as animal feed include white clover (*Trifolium repens*), sorghum (*Sorghum bicolor*), linen flax (*Linum usitatissimum*), triglochinin (*Triglochin maritima*), lima beans (*Phaseolus lunatus*), almonds (Amygdalus) and seeds of apricot (Prunus), cherries and apple (Malus). The toxic properties could be reduced by blocking the biosynthesis of cyanogenic glycosides in these plants.

The primary precursors of the naturally occuring cyanogenic glycosides are restricted to the five hydrophobic protein amino acids valine, leucine, isoleucine, phenylalanine and tyrosine and to a single non-protein amino acid, cyclopentenylglycine. These amino acids are converted in a series of reactions to cyanohydrins which are ultimately linked to a sugar residue. Amygdalin for example constitutes the O-β-gentiobioside and prunasin the O-β-glucoside of (R)-mandelonitrile. Another example of cyanogenic glycosides having aromatic aglycones is the epimeric pair of the cyanogenic glycosides dhurrin and taxiphyllin which are to be found in the genus Sorghum and Taxus, respectively. p-Hydroxymandelo-nitrile for example is converted into dhurrin by a UDPG-glycosyltransferase. Similiar glycosyltransferases are believed to be present in most plants. Vicianin and lucumin are further examples for disaccharide derivatives similiar to amygdalin. Sambunigrin contains (S)-mandelonitrile as its aglycone and is therefore epimeric to prunasin.

Examples of cyanogenic glycosides having aliphatic aglycones are linamarin and lotaustralin found in clover, linen flax, cassava and beans. A detailed review on cyanogenic glycosides and their biosynthesis can be found in Conn, Naturwissenschaften 66:28–34, 1979, herein incorporated by reference.

The biosynthetic pathway for the cyanogenic glucoside dhurrin derived from tyrosine has been extensively studied (Halkier et al, 'Cyanogenic glucosides: the biosynthetic pathway and the enzyme system involved' in: 'Cyanide compounds in biology', Wiley Chichester (Ciba Foundation Symposium 140), pages 49–66, 1988; Halkier and Moller, Plant Physiol. 90:1552–1559, 1989; Hallder et al, The J. of Biol. Chem. 264:19487–19494, 1989; Halkier and Moller, Plant Physiol. 96:10–17, 1990, Halkier and Moller, The J. of

Biol. Chem. 265:21114–21121, 1990; Halkier et al, Proc. Natl. Acad. Sci. USA 88:487–491, 1991; Sibbesen et al, in: 'Biochemistry and Biophysics of cytochrome P-450. Structure and Function, Biotechnological and Ecological Aspects', Archakov, A. I. (ed.), 1991, Koch et al, 8th Int. Conf. on Cytochrome P450, Abstract PII.053; and Sibbesen et al, 8th Int. Conf. on Cytochrome P450, Abstract PII.016). It has been found that L-Tyrosine is converted to p-hydroxymandelonitrile, the precursor of dhurrin with N-hydroxytyrosine and supposedly N,N-dihydroxytyrosine, 2-nitroso-3-(p-hydroxyphenyl)propionic acid, (E)- and (Z)-p-hydroxyphenylacetaldehydoxime, and p-hydroxyphenylacetonitrile as key intermediates. Two monooxygenases dependent on cytochrome P-450 have been reported to be involved in this pathway. A similiar pathway also involving cytochrome P-450 dependent monooxygenases has been demonstrated for the synthesis of linamarin and lotaustralin from valine and isoleucine respectively in cassava (Koch et al, Archives of Biochemistry and Biophysics, 292:141–150, 1992).

It has now surprisingly been found that the complex pathway from L-tyrosine to p-hydroxy-mandelonitrile summarized above can be reconstituted by two enzymes only, which turn out to be identical to the cytochrome P-450 dependent monooxygenases. This result is very surprising given the high degree of complexity of the pathway reflected by its numerous intermediates. Thus the two cytochrome P-450 monooxygenases are multifunctional. A first enzyme, designated P-450$_I$, converts the parent amino acid to the oxime. A second enzyme, designated P-450$_{II}$, converts the oxime to the cyanohydrin. Multifunctional cytochrome P-450 enzymes have not previously been found and described in plants.

Glucosinolates are hydrophilic, non-volatile thioglycosides found within several orders of dicotyledoneous angiosperms (Cronquist, 'The Evolution and Classification of Flowering Plants, New York Botanical Garden, Bronx, 1988). Of greatest economic significance is their presence in all members of the Brassicaceae (order of Capparales), whose many cultivars have for centuries provided mankind with a source of condiments, relishes, salad crops and vegetables as well as fodders and forage crops. More recently, rape (especially *Brassica napus* and *Brassica campestris*) has emerged as a major oil seed of commerce. About 100 different glucosinolates are known possessing the same general structure but differing in the nature of the side chain. Glucosinolates are formed from protein amino acids either directly or after a single or multiple chain extension (Underhill et al, Biochem. Soc. Symp. 38:303–326, 1973). N-hydroxy amino acids and aldoximes which have been identified as intermediates in the biosynthesis of cyanogenic glycosides also serve as efficient precursors for the biosynthesis of glucosinolates (Kindl et al, Phytochemistry 7:745–756, 1968; Matsuo et al, Phytochemistry 11:697–701, 1972; Underhill, Eur. J. Biochem. 2:61–63, 1967).

It has now surprisingly been found that the cytochrome P-450$_I$ involved in cyanogenic glycoside synthesis is very similiar to the corresponding biosynthetic enzyme in glucosinolate synthesis.

The reduction of the complex biosynthetic pathway for cyanohydrins described above to the catalytic activity of only two enzymes, cytochrome P-450$_I$ and P-450$_{II}$, allows the introduction of the biosynthetic pathway of dhurrin into plants, which plants in their wildtype phenotype do not normally produce cyanogenic glycosides. By transfection of gene constructs coding for one or both of the two cytochrome P-450 monooxygenases it will be possible to either

reconstitute or newly establish a biosynthetic pathway for cyanogenic glycosides. It is therefore an object of the present invention to provide genes coding for cytochrome P-450 monooxygenases active in the biosynthesis of cyanogenic glycosides.

The introduction of a biosynthetic pathway for cyanogenic glycosides into plants by methods known in the art, which in their wildtype phenotype do not express these glycosides is of great interest. This is due to the surprising finding of the present invention that cyanogenic glycosides can be toxic to insects, acarids, and nematodes. Therefore, the introduction or reconstitution of a biosynthetic pathway for cyanogenic glycosides in plants or certain plant tissues will allow to render plants toxic to insects, acarids or nematodes and thus help to reduce the damage to the crop by pests. In combination with other insecticidal principles such as *Bacillus thuringiensis* endotoxins the damage to the crop by pests could be even further reduced.

Alternatively, the sequences of the genes encoding the monooxygenases according to the invention can be used to design DNA plasmids which upon transfection into a plant containing cyanogenic glycosides such as cassava, sorghum or barley eliminate cyanogenic glycosides normally produced in wildtype plants. This can be achieved by expression of antisense or sense RNA or of ribozymes as described in EP-458 367 A1, EP-240 208-A2, U.S. Pat. No. 5,231,020, WO 89/05852, and WO 90/11682 which RNA inhibits the expression of monooxygenases according to the invention. This is of great interest as in spite of numerous efforts it has not been possible through traditional plant breeding to completely remove the cyanogenic glycosides from for example cassava and sorghum. On the other hand it has been shown that elevated amounts of cyanogenic glycosides in the epidermal cells of barley cultivars confer increased sensitivity to attack by the mildew fungus *Erysiphe graminis* (Pourmohensi, PhD thesis, Göttingen, 1989; Ibenthal et al, Angew. Bot. 67:97–106, 1993). A similiar effect has been observed in the cyanogenic rubber tree *Hevea brasiliensis* upon attack by the fungus *Microcyclus ulei* (Lieberei et al, Plant Phys. 90:3–36, 1989) and with flax attacked by *Colletotrichum lini* (Lüdtke et al, Biochem. Z. 324:433442, 1953). In these instances the quantitative resistance of the plants stipulated above and of other plants, where cyanogenic glycosides confer increased sensitivity to attack by microorganisms, can be increased by preventing the production of cyanogenic glycosides in such plants. In barley, the cyanogenic glycosides are located in the epidermal cells. The antisene, sense or ribozyme constructs are therefore preferably but but not necessarily combined with an epidermis specific promoter.

The presence of even minor amounts of cyanogenic glycosides in plants may also cause nutritional problems due to generation of unwanted carcinogens as demonstrated in barley. Barley malt for example contains low amounts of the cyanogenic glucoside epiheterodendrin which in the cause of production of grain-based spirits can be converted to ethylcarbamate which is considered to be a carcinogen. Attempts are being made to introduce mandatory maximum allowable concentrations of ethylcarbarnate in fermented food, beverages and spirits (Food Chemical News 29:33.35, 1988).

Plants containing cyanogenic glycosides typically contain only a single cyanogenic glycoside or just a few. In certain cases it is of interest to alter the cyanogenic glycoside profile of a plant. Since cytochrome P-450$_{II}$ shows broad substrate specificity this enzyme typically converts the aldoxime produced by cytochrome P-450$_I$ into the corresponding cyanohydrin. Alteration of the chemical identity of cyanogenic glycosides produced in a specific plant can thus be accomplished by transforming a plant with an additional gene encoding an expressible cytochrome P-450$_I$ monooxygenase with a substrate specificity different from the naturally occuring enzyme.

The present invention relates primarily to a DNA molecule coding for a cytochrome P-450 monooxygenase, which catalyzes the conversion of an amino acid to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid Preferably the inventive monooxygenase catalyzes the conversion of an amino acid selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine and cyclopentenylglycine or an amino acid selected from the group consisting of L-tyrosine, L-valine and L-isoleucine. Additionally the present invention relates to a DNA molecule coding for a cytochrome P-450 monooxygenase, which monooxygenase catalyzes the conversion of said oxime to a nitrile and the conversion of said nitrile to the corresponding cyanohydrin. The DNA molecules according to the invention either correspond to naturally occuring genes or to functional homologues thereof which are the result of mutation, deletion, truncation, etc. but still encode cytochrome P-450 monooxygenases, which either catalyze the conversion of an amino acid to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid, or the conversion of said oxime to a nitrile and the subsequent conversion of said nitrile to the corresponding cyanohydrin. Both monooxygenases are able to catalyze more than one reaction of the biosynthetic pathway of cyanogenic glycosides but preferably contain a single catalytic center. The monooxygenase cytochrome P-450, converting the parent amino acid is also involved in glucosinolate biosynthesis. Because cytochrome P-450$_I$ determines the substrate specificity and thus the type of glucosinolates produced and because cytochrome P-450$_I$ constitutes the rate limiting step, the principles already described above for cyanogenic glucosides can also be used to down- or up-regulate the synthesis of glucosinolates in glucosinolate producing plants and to alter the compositon of glucosinolates produced.

The inventive DNA molecule encoding cytochrome P-450$_I$ is obtainable from plants which produce cyanogenic glycosides and glucosinolates. These plants include but are not limited to plants selected from the group consisting of the species Sorghum, Trifoliun, Linum, Taxus, Triglochin, Mannihot, Amygdalus and Prunus as well as cruciferous plants. In a preferred embodiment of the invention the DNA molecule is obtained from *Sorghum bicolor* (L.) Moench or *Manihot esculenta* Crantz. The sequence similarity between cytochrome P-450$_I$ monooxygenases from different plants producing cyanogenic glycosides or glucosinolates is evidenced by the specific cross-reactivity of antibodies prepared against cytochrome P-450$_{TYR}$ isolated from sorghum, with the corresponding cytochrome P-450 enzyme in cassava and with the corresponding enzyme in the glucosinolate producing plant *Tropaeolum majus*. Southern blotting using the cDNA clone encoding cytochrome P-450$_{TYR}$ shows specific and strong hybridization to genomic DNA isolated from cassava, *Tropaeolum majus*, and rape. Of all approximately 250 known published sequences for cytochrome P-450 enzymes, cytochrome P-450$_{TYR}$ shows the highest sequence similarity to the petunia 3'5'-flavonoid hydroxylase (30.8%) and 28% sequence similiarity to CYP1A2 from rabbit. The group of cytochrome P-450$_I$ monooxygenases functionally characterized by catalyzing the conversion of

an amino acid to the corresponding aldoxime can thus be defined as cytochrome P-450 enzymes the amino acid sequence of which exhibits a 32% or higher sequence similarity and preferably a 40% or higher sequence similarity to that of cytochrome P-450$_{TYR}$ Cytochrome P-450 gene protein families are defined as having less than 40% amino acid identity to a cytochrome P-450 protein from any other family. Consequently, cytochrome P-450$_{TYR}$ belongs to a new P-450 protein family.

The inventive DNA molecule encoding cytochrome P-450$_{II}$ is obtainable from plants which produce cyanogenic glycosides. In a preferred embodiment of the invention the DNA molecule is obtained from Sorghum bicolor (L.) Moench or Manihot esculenta Crantz. The enzyme isolated from Sorghum bicolor (L.) Moench is designated cytochrome P-450$_{Ox}$. The catalytic properties of this enzyme resembles those of a cytochrome P-450 activity reported in microsomes from rat liver (DeMaster et al, J. Org. Chem. 5074–5075, 1992) which has neither been isolated nor further characterized. A characteristic of cytochrome P-450$_{Ox}$ and of other members belonging to the cytochrome P-450$_{Ox}$ family is that dehydration of the oxime to the corresponding nitrile is dependent on the presence of NADPH but that this dependence can be overcome by the addition of sodium dithionite or other reductants. Cytochrome P-450 enzymes able to convert aldoximes into cyanohydrins might be present in most living organisms.

For the purposes of gene manipulation using recombinant DNA technology the DNA molecule according to the invention may in addition to the gene coding for the monooxygenase comprise DNA which allows for example replication and selection of the inventive DNA in microorganisms such as *E. coli,* Bacillus, Agrobacterium, Streptomyces or yeast. It may also comprise DNA which allows the monooxygenase genes to be expressed and selected in homologous or heterologous plants. Such sequences comprise but are not limited to genes whose codon usage has been adapted to the codon usage of the heterologous plant as described in WO 93/07278; to genes conferring resistance to neomycin, kanamycin, methotrexate, hygromycin, bleomycin, streptomycin, or gentamycin, to aminoethylcystein, glyophosphate, sulfonylurea, or phosphinotricin; to scorable marker genes such as galactosidase; to its natural promoter and transcription termination signals; to promoter elements such as the 35S and 19S CaMV promoters, or tissue specific plant promoters such as promoters specific for root (described for example in EP-452 269-A2, WO 91/13992, U.S. Pat. No. 5,023,179), green leaves such as the maize phosphoenol pyruvate carboxylase (PEPC), pith or pollen (described for example in WO 93/07278), or inducible plant promoters (EP 332 104); and to heterologous transcription termination signals.

The present invention also relates to monooxygenases which catalyze the conversion of an amino acid preferably selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine and cyclopentenylglycine to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid (cytochrome P-450$_I$); or the conversion of said oxime to a nitrile and the conversion of said nitrile to the corresponding cyanohydrine (cytochrome P-450$_{II}$). In a preferred embodiment of the invention the monooxygenases are purified and can be used to establish monoclonal or polyclonal antibodies which specifically bind to the monooxygenases.

In another preferred embodiment of the invention the cytochrome P-450$_{II}$ monooxygenase can be isolated from Sorghum, has a molecular weight of 51 kD as determined by SDS-PAGE and comprises the N-terminal sequence MDLADIPKQQRLMAGNALVV (SEQ ID NO: 12).

For other cytochrome P-450$_{II}$ enzymes, the N-terminal sequences may be different.

Optionally, a P-450$_{II}$ monooxygenase might also comprise one of the following sequences:

--ARLAEIFATII-- (SEQ ID NO: 13)

--EDFTVTTK-- (SEQ ID NO: 14)

--QYAALGSVFTVPII-- (SEQ ID NO: 15)

--XXPFPI-- (SEQ ID NO: 16).

Another embodiment of the present invention deals with a method for the preparation of cDNA coding for a cytochrome P-450 monooxygenase, which either catalyzes the conversion of an amino acid preferably selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine and cyclopentenylglycine, to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid (cytochrome P-450$_I$); or the conversion of said oxime to a nitrile and the conversion of said nitrile to the corresponding cyanohydrin (cytochrome P-450$_{II}$); comprising

(a) isolating and solubilizing microsomes from plant tissue producing cyanogenic glycosides or glucosinolates,

(b) purifying the cytochrome P-450 monooxygenase,

(c) raising antibodies against the purified monooxygenase,

(d) probing a cDNA expression library of plant tissue producing cyanogenic glycosides or glycosinolates with said antibody, and

(e) isolating clones which express the monooxygenase.

Microsomes can be isolated from plant tissues which show a high activity of the enzyme system responsible for biosynthesis of the cyanogenic glycosides. These tissues may be different from plant species to plant species. A preferred source of microsomes are freshly isolated shoots harvested 1 to 20 days, preferably 2 to 10 days and most preferably 2 to 4 days after germination. Etiolated seedlings are preferred from plant producing cyanogenic glycosides but light grown seedlings may also be used. Following isolation the microsomes are solubilized in buffer containing one or more detergents. Preferred detergents are RENEX 690 (J. Lorentzen A/S, Kvistgard, Denmark), reduced Triton X-100 (RTX-100) and CHAPS.

The cytochrome P-450 monooxygenases can be purified applying standard techniques for protein purification such as ultracentrifugation, fractionated precipitation, dialysis, SDS-PAGE and column chromatography. Possible columns comprise but are not limited to ion exchange columns such as DEAE Sepharose, Reactive dye columns such as Cibacron yellow 3 agarose, Cibacron blue agarose and Reactive red 120 agarose, and gel filtration columns such as Sephacryl S-1000. The cytochrome P-450 content of the individual fractions can be determined from carbon monoxide difference spectra.

The purified proteins can be used to elicit antibodies in for example mice, goats, sheeps, rabbits or chickens upon injection. 5 to 50 μg of protein are injected several times during approximately 14 day intervals. In a preferred embodiment of the invention 10 to 20 μg are injected 2 to 6 times in 14 day intervals. Injections can be done in the presence or absence of adjuvants. Immunoglobulins are purified from the antisera and spleens can be used for hybridoma fusion as described in Harlow and Lane, 'Antibodies: A Laboratory Manual', Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1988, herein incorporated by reference. Antibodies specifically binding to a cytochrome P-450 monooxygenase can also be used in plant breeding to detect plants producing altered amounts of cytochrome P-450 monooxygenases and thus altered amounts of cyanogenic glycosides.

The methods for the preparation of plant tissue cDNA libraries are extensively described in Sambrook et al, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, the essential parts of which regarding preparation of cDNA libraries are herein incorporated by reference. PolyA$^+$ RNA is isolated from plant tissue which shows a high activity of the enzyme system responsible for biosynthesis of the cyanogenic glycosides or glucosinolates. These tissues may be different from plant species to plant species. A preferred tissue for polyA$^+$ RNA isolation is the tissue of freshly isolated shoots harvested 1 to 20 days, preferably 2 to 10 days and most preferably 2 to 4 days after germination. When cDNA libraries are made from glucosinolate producing plants older or mature plant tissue may also be used. The obtained cDNA libraries can be probed with antibodies specifically binding the cytochrome P-450 monooxygenase and clones expressing the monooxygenase can be isolated.

An alternative method for the preparation of cDNA coding for a cytochrome P-450 monooxygenase comprises

(a) isolating and solubilizing microsomes from plant tissue producing cyanogenic glycosides or glucosinolates,
(b) purifying the cytochrome P-450 monooxygenase,
(c) obtaining a complete or partial protein sequence of the monoxygenase,
(d) designing oligonucleotides specifying DNA coding for 4 to 15 amino acids of said monooxygenase protein sequence
(e) probing a cDNA library of plant tissue producing cyanogenic glycosides or glucosinolates with said oligonucleotides, or DNA molecules obtained from PCR amplification of cDNA using said oligonucleotides, and
(f) isolating clones which encode cytochrome P-450 monooxygenase.

Amino acid sequences of internal peptides which are the result of protease digestion can be obtained by standard techniques such as Edman degradation. Oligonucleotides specifying DNA coding for partial protein sequences of the inventive monooxygenases are obtained by reverse translation of parts of the protein sequence according to the genetic code. Protein sequences encoded by DNA sequences of low degeneracy are preferred for reverse translation. Their length ranges from 4 to 15 and preferably from 5 to 10 amino acids. If necessary the codons used in the oligonucleotides can be adapted to the codon usage of the plant source (Murray et al, Nucleic Acids Research 17:477–498, 1989). The obtained oligonucleotides can be used to probe cDNA libraries as described in Sambrook et al, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, for clones which are able to basepair with said oligonucleotides. Alternatively, oligonucleotides can be used in a polymerase chain reaction, the methodology of which is known in the art, with plant cDNA as the template for amplification. In this case the obtained amplification products are used to probe the cDNA libraries. Clones encoding cytochrome P-450 monooxygenases are isolated.

An alternative method of cloning genes is based on the construction of a gene library composed of expression vectors. In that method, analogously to the methods already described above, genomic DNA, but preferably cDNA, is first isolated from a cell or a tissue capable of expressing a desired gene product—in the present case a P-450 monooxygenase—and is then spliced into a suitable expression vector. The gene libraries so produced can then be screened using suitable measures, preferably using antibodies, and those clones selected which comprise the desired gene or at least part of that gene as an insert.

Alternatively, total DNA from the DNA library, preferably from the cDNA library, can be prepared and used as a template for a PCR reaction with primers representing low degeneracy portions of the amino acid sequence. Preferably, the primers used will generate PCR products that represent a significant portion of the nucleotide sequence. The PCR products can be further probed to determine if they correspond to a portion of the P-450 monooxygenase gene using a synthetic oligonucleotide probe corresponding to an amino acid fragment sequence located in the interior or middle region of the P-450 monooxygenase protein.

The cDNA clones and PCR products prepared as described above or fragments thereof may be used as a hybridization probe in a process of identifying further DNA sequences from a homologous or a heterologous source organism encoding a protein product that exhibits P-450 monooxygenase activity such as, for example, a fungi or a heterologous plant. A suitable source would be tissue from plants containing cyanogenic glycosides or glucosinolates.

They may also be used as an RFLP marker to determine, for example, the location of the cytochrome P-450 monooxygenase gene or a closely linked trait in the plant genome or for marker assisted breeding [EP-A 306,139; WO 89/07647].

Using the methods described above it is thus possible to isolate a gene that codes for a P-450 monooxygenase.

Genes encoding cytochrome P-450 monooxygenase can be used in a method for producing a purified recombinant cytochrome P-450 monooxygenase which monooxygenase either catalyzes the conversion of an amino acid preferably selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine and cyclopentenylglycine to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid; or the conversion of said oxime to a nitrile and the conversion of said nitrile to the corresponding cyanohydrine; comprising

(a) engineering the gene encoding said monooxygenase to be expressible in a host organism such as bacteria, yeast or insect cells,
(b) transforming said host organism with the engineered gene, and
(c) isolating the protein from the host organism or the culture supernatant.

In a preferred embodiment of the invention the method is used to obtain purified recombinant cytochrome P-450$_{TYR}$, P-450$_{Ox}$ or cytochrome P-450$_{TYR}$ which has been modified by known techniques of gene technology. Preferably the modifications lead to increased expression of the recombinant protein or to altered substrate specificity.

The inventive DNA molecules can be used to obtain transgenic plants resistant to insects or acarids examples of which are listed but not limited to those in Table B as well as nematodes. Preferably the transgenic plants are resistant to Coleoptera and Lepidoptera such as western corn root worm (*Diabrotica virgifera virgifera*), northern corn root worm (*Diabrotica longicornis barberi*), southern corn rootworm (*Diabrotica undecimpunctata howardi*), cotton bollworm, European corn borer, corn root webworm, pink bollworm and tobacco budworm. The transgenic plants comprise DNA coding for monooxygenases which catalyze the conversion of an amino acid to the corresponding N-hydroxyamino acid and the oxime derived from this N-hydroxyamino acid; or the conversion of said oxime to a nitrile and the conversion of said nitrile to the corresponding cyanohydrine. In addition the transgenic plants may comprise monooxygenase genes genetically linked to herbicide

resistance genes. The transgenic plants are preferably monocotyledoneous or dicotyledoneous plants examples of which ar listed in Table A. Preferably they are selected from the group consisting of maize, rice, wheat, barley, sorghum, cotton, soybeans, sunflower, grasses and oil seed rape. The plants can be obtained by a method comprising

(a) introducing into a plant cell or plant tissue which can be regenerated to a complete plant, DNA comprising a gene expressible in that plant encoding an inventive monooxygenase,

(b) selecting transgenic plants, and

(c) identifying plants which are resistant to insects, acarids, or nematodes.

The inventive DNA molecules can also be used to obtain transgenic plants expressing anti-sense or sense RNA or ribozymes targeted to the genes of the endogenous P-450 monooxygenases the expression of which reduces the expression of cytochrome P-450 monooxygenases. Such plants show improved disease resistance or nutritive value due to reduced expression of cyanogenic glycosides or glucosinolates. The plants can be obtained with a method comprising

(a) introducing into a plant cell or tissue which can be regenerated to a complete plant, DNA encoding sense RNA, anti sense RNA or a ribozyme, the expression of which reduces the expression of cytochrome P-450 monooxygenases according to claims 1 or 8,

(b) selecting transgenic plants, and

(c) identifying plants with improved disease resistance or nutritive value.

A number of very efficient processes are available for introducing DNA into plant cells, which processes are based on the use of gene transfer vectors or on direct gene transfer processes.

One possible method of inserting a gene construct into a cell makes use of the infection of the plant cell with *Agrobacterium tumefaciens* and/or *Agrobacterium rhizogenes,* which has been transformed with the said gene construction. The transgenic plant cells are then cultured under suitable culture conditions known to the person skilled in the art, so that they form shoots and roots and whole plants are finally formed.

Within the scope of this invention is the so-called leaf disk transformation using Agrobacterium (Horsch et al, Science 227:1229–1231, 1985) can also be used. Sterile leaf disks from a suitable target plant are incubated with Agrobacterium cells comprising one of the chimaeric gene constructions according to the invention, and are then transferred into or onto a suitable nutrient medium. Especially suitable, and therefore preferred within the scope of this invention, are LS media that have been solidified by the addition of agar and enriched with one or more of the plant growth regulators customarily used, especially those selected from the group of the auxins consisting of a-naphthylacetic acid, picloram, 2,4,5-trichlorophenoxyacetic acid, 2,4-dichlorophenoxyacetic acid, indole-3-butyric acid, indole-3-lactic acid, indole-3-succinic acid, indole-3-acetic acid and p-chlorophenoxyacetic acid, and from the group of the cytokinins consisting of kinetin, 6-benzyladenine, 2-isopentenyladenine and zeatin. The preferred concentration of auxins and cytokinins is in the range of 0.1 mg/l to 10 mg/l.

After incubation for several days, but preferably after incubation for 2 to 3 days at a temperature of 20° C. to 40° C., preferably from 23° C. to 35° C. and more preferably at 25° C. and in diffuse light, the leaf disks are transferred to a suitable medium for the purpose of shoot induction. Especially preferred for the selection of the transformants is an LS medium that does not contain auxin but contains cytokinin instead, and to which a selective substance has been added. The cultures are kept in the light and are transferred to fresh medium at suitable intervals, but preferably at intervals of one week. Developing green shoots are cut out and cultured further in a medium that induces the shoots to form roots. Especially preferred within the scope of this invention is an LS medium that does not contain auxin or cytokinin but to which a selective substance has been added for the selection of the transformants.

In addition to Agrobacterium-mediated transformation, within the scope of this invention it is possible to use direct transformation methods for the insertion of the gene constructions according to the invention into plant material.

For example, the genetic material contained in a vector can be inserted directly into a plant cell, for example using purely physical procedures, for example by microinjection using finely drawn micropipettes (Neuhaus et al, Theoretical and Applied Genetics 74:363–373, 1987), electroporation (D'Halluin et al, The Plant Cell 4:1495–1505, 1992; WO 92/09696), or preferably by bombarding the cells with microprojectiles that are coated with the transforming DNA ("Microprojectile Bombardment"; Wang et al, Plant Molecular Biology 11:433–439, 1988; Gordon-Kamm et al, The Plant Cell 2:603–618, 1990; McCabe et al, Bio/Technology 11:596–598, 1993; Christou et, Plant Physiol. 87:671–674, 1988; Koziel et al, Biotechnology 11: 194–200, 1993). Moreover, the plant material to be transformed can optionally be pretreated with an osmotically active substance such as sucrose, sorbitol, polyethylene glycol, glucose or mannitol.

Other possible methods for the direct transfer of genetic material into a plant cell comprise the treatment of protoplasts using procedures that modify the plasma membrane, for example polyethylene glycol treatment, heat shock treatment or electroporation, or a combination of those procedures (Shillito et al, Biotechnology 3:1099–1103, 1985).

A further method for the direct introduction of genetic material into plant cells, which is based on purely chemical procedures and which enables the transformation to be carried out very efficiently and rapidly, is described in Negrutiu et al, Plant Molecular Biology 8:363–373, 1987.

Also suitable for the transformation of plant material is direct gene transfer using co-transformation (Schocher et al, Bio/Technology 4:1093–1096, 1986).

The list of possible transformation methods given above by way of example does not claim to be complete and is not intended to limit the subject of the invention in any way.

In another embodiment of the invention target plants are exposed to a pesticidally effective amount of a cyanogenic glycoside to control insects, acarids, or nematodes attacking a monocotyledonous or dicotyledonous plant selected from the group of plant types consisting of Cereals, Protein Crops, Fruit Crops, Vegetables and Tubers, Nuts, Oil Crops, Sugar Crops, Forage and Turf Grasses, Forage Legumes, Fiber Plants and Woody Plants, Drug Crops and Spices and Flavorings.

The following examples further describe the materials and methods used in carrying out the invention and the subsequent results. They are offered by way of illustration, and their recitation should not be considered as a limitation of the claimed invention.

EXAMPLES

Example 1

Preparation of microsomes

Seeds of Sorghum bicolor (L.) Moench (hybrid S-1000) are obtained from Seedtec International Inc. (Hereford, Tex.) and germinated in the dark for 2 days at 28° C. on metal screens covered with gauze. Transfer of the seeds to germination trays is carried out under dim green light. Microsomes are prepared from approximately 3 cm tall etiolated seedlings. The seedlings are harvested and homogenized using a mortar and pestle in 2 volumes (v/w) of 250 mM sucrose, 100 mM tricine (pH 7,9), 50 mM NaCl, 2 mM EDTA and 2 mM DTT. Polyvinylpolypyrrolidone is added (0.1 g/g fresh weight) prior to homogenization. The homogenate is filtered through 22 μm nylon cloth and centrifuged 20 minutes at 48000 g. The supernatant is centrifuged for 1 hour at 165000 g. The microsomal pellet is resuspended and homogenized in isolation buffer using a Potter-Elvehjem homogenizer fitted with a teflon pestle. After recentrifugation and rehomogenization, the homogenate is dialyzed overnight against 50 mM Tricine (pH 7,9), 2mM DTT under a nitrogen atmosphere.

Example 2

Enzyme assays: Determination of total cytochrome P-450

Quantitative determination of total cytochrome P-450 is carried out by difference spectroscopy using an extinction difference coefficient of 91 $mM^{-1}cm^{-1}$ for the complex between reduced cytochrome P-450 and carbon monoxide ($A_{450-490}$) (Omura et al, J. Biol. Chem. 239:2370–2378, 1964). Cytochrome P-450 substrate binding spectra are recorded with stepwise increased substrate concentration until saturating conditions are reached.

Example 3

Purification of cytochrome P-450$_{TYR}$ and P-450$_{Ox}$

Buffer A:	Buffer B:	Buffer C:
8.6% glycerol	8.6% glycerol	8.6% glycerol
10 mM KH_2PO_4/ K_2HPO_4 (pH 7.9)	40 mM KH_2PO_4/ K_2HPO_4 (pH 7.9)	40 mM KH_2PO_4/ K_2HPO_4 (pH 7.9)
0.20 mM EDTA	5.0 mM EDTA	5.0 mM EDTA
2.0 mM DTT	2.0 mM DTT	2.0 mM DTT
1.0% RENEX 690	1.0% RENEX 690	1.0% CHAPS
0.05% RTX-100	0.05% RTX-100 0.2% CHAPS	0.05% RTX-100

Buffers are degassed three times by stirring in vacuo before detergent and DTT are added. Between each degassing, the buffer is flushed with argon.

Microsomes (400 mg protein in 20 ml) are diluted to 100 ml with a buffer composed of 8.6% glycerol, 10 mM KH_2PO_4/K_2HPO_4 (pH 7.9). The microsomes are solubilized by slow addition of 100 ml of the same buffer containing 2% RENEX 690 and 0.2% RTX-100 and constant stirring for 30 minutes. Solubilized cytochrome P-450 is obtained as the supernatant after centrifugation for 30 minutes at 200 000 g in a Beckman 70:Ti rotor. The supernatant (190 ml) is applied (flow rate 100 ml/h) to a column (5×5 cm) of DEAE Sepharose fast flow/S-100 Sepharose (20:80 wet volumes) equilibrated in buffer A. The ion exchange resin DEAE-Sepharose is diluted with the gel filtration material Sephacryl S-100 in the ratio 1:4 to avoid too high concentrations of the cytochrome P-450 enzymes upon binding, which sometimes results in irreversible aggregation. The column is washed with 150 ml buffer A after which the total amount of cytochromes P-450 including cytochrome P-450$_{TYR}$ and cytochrome P-450$_{Ox}$ is eluted with buffer B in a total volume of 150 ml. During this procedure, NADPH-cytochrome P-450-oxidoreductase and Cytochrome b_5 remain bound to the column and may subsequently be eluted and separated with buffer B and a gradient of 0–300 mM KCl.

The cytochrome P-450 eluate is adjusted to 1.0% CHAPS, stirred for 30 minutes and then directly applied to a 25 ml (2.6×5 cm) column of Reactive yellow 3 sepharose equilibrated in buffer C+1,0% RENEX 690. The flow rate used is 25 ml/h. The column is washed with buffer C until the absorbance A_{280} shows that RENEX 690 is washed out. Cytochrome P-450$_{TYR}$ does not bind to this column, and is obtained in the run-off and wash. Subsequently the column is eluted with 400 mM KCl in buffer C. The cytochrome P-450$_{Ox}$ containing fractions are combined yielding approximately 60 ml and diluted with 5 volumes of buffer C to lower the KCl strength and permit rebinding of cytochrome P-450$_{Ox}$ on a second Reactive yellow 3 column. This column is eluted with a KCl gradient (0–500 mM) in a total volume of 100 ml in buffer C. This serves to elute cytochrome P-450$_{Ox}$.

The cytochrome P-450$_{Ox}$ pool from the yellow agarose is diluted 5 times with buffer C to 20–25 mM KCl and applied to a Cibachron blue agarose column (0.9×6 cm) equilibrated in buffer C. The flow rate used is 8 ml/h. The column is washed with 20 ml buffer C at the same flow rate. Cytochrome P-450$_{Ox}$ is eluted with a gradient of KCl, 0–2.0 M in buffer C in a total volume of 30 ml.

The runoff from the first yellow 3 agarose column is applied (flow rate 40 ml/h) to a column (2.8×8 cm) of Cibachron Blue Agarose equilibrated in buffer C. The column is subsequently washed with buffer C and the cytochrome P-450$_{TYR}$ is eluted with a 0–500 mM linear KCl-gradient (2×100 ml) in buffer C. The combined cytochrome P-450 fractions are diluted 5 times with buffer C and applied (flow rate 7 ml/h) to a column (0.9×5 cm) of Reactive red 120 agarose equilibrated in buffer C. The column is washed with 25 ml buffer C and cytochrome P-450$_{TYR}$ is eluted with a 0–1.0 M KCl linear gradient (2×30 ml) in buffer C. Optionally the eluate is gelfiltrated through a Sephadex G-50 column, equilibrated in a buffer composed of 50 mM potassium phosphate (pH 7.9)/400 mM KCl/0.1% CHAPS/2 mM DTT. The eluted cytochrome P-450$_{TYR}$ is dialyzed for 2 hours against 50 mM potassium phosphate (pH 7.9)/2 mM DTT, diluted 4 fold with dialysis buffer in an Amicon ultrafiltration cell fitted with a YM-30 membrane and concentrated to 1.45 nmols/ml.

All procedures are carried out at 4° C. The total cytochrome P-450 content of the individual fractions is determined from the carbon monoxide difference spectrum. The absorption spectrum of the oxidized cytochrome P-450 is also recorded. The presence of a specific cytochrome P-450 is monitored by substrate binding spectra.

Example 4

Antibody preparation

Polyclonal antibodies are elicited in rabbits by six repeated subcutaneous injections (approx. 15 μg protein per rabbit per injection) at 15 day intervals of cytochrome P-450$_{TYR}$ or P-450$_{Ox}$ isolated by dye column chromatography or denatured enzyme purified by preparative SDS-PAGE. Freunds complete adjuvant is included in the first injection whereas Freunds incomplete adjuvant is used in subsequent injections. The immunoglobulin fractions of the antisera are purified by ammonium sulfate precipitation

(Harboe et al, 'A Manual of Quantitative Immunoelectrophoresis: Methods and Applications', Universitetsforlaget, Oslo, 1973). The antiobides are monospecific as demonstrated by Western blotting.

Example 5

Characterization of P-450$_{TYR}$

5.1. Substrate binding spectra of P-450$_{TYR}$

Cytochrome P-450$_{TYR}$ in the oxidized state has a strong absorption peak at 420 nm, representing the low spin state of the iron of the heme group. The binding of a ligand to the heme group shifts the absorption maximum by changing the spin state of the iron. Binding of tyrosine at the catalytic site of cytochrome P-450$_{TYR}$ induces a change of the spin state of the oxidized iron from low to high spin, and thereby changes the absorption maximum from 420 nm to 390 nm producing a type I spectrum (Jefcoate, Methods in Enzymology 52:258–279, 1978). The following experimental procedure is used to obtain the substrate binding spectrum: two identical cuvettes containing a buffered solution of the isolated cytochrome P-450 are prepared. The substrate of the enzyme is added to the sample cuvette whereas the same volume of buffer is added to the other cuvette. The difference spectrum is then recorded in an SLM-Aminco DW2c spectrophotometer. The absorption difference, $A_{390-420}$, is proportional to the concentration of cytochrome P-450$_{TYR}$ with a bound substrate at its active site. If a saturating concentration of substrate is added to the sample cuvette, the absorption difference is proportional to the concentration of the substrate specific cytochrome P-450 in the cuvettes. The saturating concentration of the substrate is determined by titrating the cytochrome P-450 sample with increasing amounts of substrate and monitoring $A_{390-420}$.

If a cytochrome P-450 sample can be saturated with two different substrates, there may be two different cytochrome P-450 enzymes in the sample, or there may be one cytochrome P-450 enzyme able to bind to both substrates. To discriminate between these possibilities, saturating amounts of the two substrates are added sequentially and the $A_{390-420}$ absorption change is monitored. If, independent of the order of addition, the addition of the second sample gives rise to an increased $A_{390-420}$ value compared to the value after the addition of the first substrate, the two substrates are bound by different enzymes. If $A_{390-420}$ remains unchanged upon addition of the second substrate, independent of the order of addition, both substrates bind to the same active site, i.e. to the same cytochrome P-450 enzyme. The data shown in Tables C and D below represent results of a typical experiment.

TABLE C

To 500 µl of isolated cytochrome P-450$_{TYR}$ dissolved in 50 mM Tricine pH 7,9 tyrosine is added until saturation concentration is reached followed by addition of N-hydroxytyrosine:

Added substrate	initial $A_{390-420}$	dilution factor	resulting $A_{390-420}$
30 µl 5 mM tyrosine	0,0437	530/500	0,0463
60 µl 5 mM tyrosine	0,0496	560/500	0,0556
90 µl 5 mM tyrosine	0,0486	590/500	0,0573
+100 µl 20 mM N-hydroxytyrosine	0,0409	690/500	0,0564

TABLE D

Addtion of N-hydroxytyrosine until saturation concentration is reached, followed by addition of tyrosine

Added substrate	initial $A_{390-420}$	dilution factor	resulting $A_{390-420}$
50 µl 20 mM N-hydroxytyrosine	0,0689	550/500	0,0758
120 µl 5 mM N-hydroxytyrosine	0,0919	620/500	0,1140
140 µl 5 mM N-hydroxytyrosine	0,0911	640/500	0,1166
+90 µl 5 mM tyrosine	0,0726	730/500	0,1060

Both tyrosine and N-hydroxytyrosine produce a type I binding spectrum. The data show, that tyrosine and N-hydroxytyrosine bind to the same active site, that is the same cytochrome P-450, thus demonstrating that cytochrome P-450$_{TYR}$ is multifunctional. From the amounts of cytochrome P-450$_{TYR}$ used the absorption coefficient ($\epsilon_{390-420}$) is calculated to be 67 $cm^{-1}mM^{-1}$. A complete transition from a low spin state to a high spin state would have resulted in an absorption coefficient of 138 $cm^{-1}mM^{-1}$.

5.2. Molecular weight and Amino acid sequence data

The molecular weight of P-450$_{TYR}$ as determined by SDS-PAGE is 57 kD.

Amino acid sequences are obtained by automated Edman degradation. The internal polypeptides are obtained by trypsin digestion of the purified protein and subsequent separation of the peptides using reverse phase HPLC.

N-terminal sequence:

--MATMEVEAAAATVLAAP-- (SEQ ID NO: 3)

Internal sequences:

--VWDEPLR-- (SEQ ID NO: 4)
--YVYNLATK-- (SEQ ID NO: 5)
--SDTFMATPLVSSAEPR-- (SEQ ID NO: 6)
--AQSQDITFAAVDNPSNAVEXALAEMVNNPEVMAK-- (SEQ ID NO: 7)
--AQGNPLLTIEEVK-- (SEQ ID NO: 8)
--LVQESDIPK-- (SEQ ID NO: 9)
--ISFSTG-- (SEQ ID NO: 10)
--LPAHLYPSISLH-- (SEQ ID NO: 11)

5.3. Reconstitution of cytochrome P-450$_{TYR}$ activity:

Reconstitution of the enzyme activity of a microsomal P-450 enzyme is accomplished by insertion of the cytochrome P-450 enzyme and the corresponding NADPH-cytochrome P-450 oxidoreductase into appropriate lipid micelles made from different commercially available lipids. Isolation of NADPH-cytochrome P-450 oxidoreductase is done according to Hallder and Moller, Plant Physiol. 96:10–17, 1990. A mixture of lipids can be used but with cytochrome P-450$_{TYR}$ di-lauroyl-phosphatidyl choline (DLPC) provides the best enzymatic activity. One rate limiting factor of this rate limiting reaction is the number of correctly formed complexes of cytochrome P-450$_{TYR}$ and NADPH-cytochrome P-450 oxidoreductase. Excess amounts of the oxidoreductase and concentrated enzyme solutions ensure a sufficient number of active complexes.

A reconstituted enzyme is obtained using the following components:

Cytochrome P-450$_{TYR}$:	100 µg/ml in 50 mM potassium phosphate buffer pH 7,9
Oxidoreductase, purified from Sorghum bicolor:	100 µg/ml in 50 mM potassium phosphate buffer pH 7,9

-continued

Lipid:	10 mg/ml di-lauroyl-phosphatidyl choline, sonicated in 50 mM potassium phosphate buffer pH 7,9
NADPH:	25 mg/ml H_2O
^{14}C-tyrosine:	commercially available from Amersham; [U—^{14}C]—L-tyrosine 0.5 μCi

10 μg of the lipid suspension is mixed in a glass vial with 50 μl of the cytochrome P-450$_{TYR}$ (0–1.5 pmol) solution. 50 μl of the oxidoreductase (0–0.15 U) solution is added and then 10 μl tyrosine solution and 10 μl NADPH solution are added and the mixture is sonicated in a Branson 5200 sonication bath for one minute. The reaction mixture is subsequently incubated for 1 hour at 30° C. At the end of the incubation period the reaction is stopped by transferring the glass vials onto ice. Radioactively labelled intermediates formed are extracted into 50 μl ehtyl acetate, applied to a silica coated TLC plate and developed using an ethyl acetate/toluene (1:5 v/v) mixture as mobile phase. The resultant product, p-hydroxyphenylacetaldehyde oxime is visualized by autoradiography of the TLC plate. Alternatively, the intermediates are analyzed by reverse-phase HPLC coupled to a Berthold radioactiviy monitor. The HPLC separation was carried out using a nucleosil 100-10C$_{18}$ column isocratically eluted with 1.5% 2-propanol in 25 mM Hepes pH 7.9 (Halkier et al, J. Biol. Chem. 264:19487–19494, 1989). Control samples may be made by omitting either cytochrome P-450 oxidoreductase or NADPH.

When reconstituted into micelles cytochrome P-450$_{TYR}$ catalyzes the conversion of L-tyrosine all the way to p-hydroxyphenyl-acetaldehyde oxime. The K_m and turn-over number of the enzyme are 0.14 mM and 198 mind, respectively, when assayed in the presence of 15 mM NaCl, whereas the values are 0.21 mM and 228 min^{-1} when assayed in the absence of added salt.

The formation of p-hydroxyphenyl-acetaldehyde oxime demonstrates that cytochrome P-450$_{TYR}$ is a multi-functional heme-thiolate protein catalyzing reactions in addition to the initial N-hydroxylation of L-tyrosine. The E/Z ration of the parahydroxyphenyl-acetaldehyde oxime produced by the reconstituted cytochrome P-450$_{TYR}$ and determined by HPLC chromatography is 69/31. Using the TLC/autoradiography system, minute amounts of radiolabelled products comigrating with authentic p-hydroxybenzaldehyde and 1-nitro-2(p-hydroxyphenyl) ethane are detected in the reaction mixtures.

5.4. Inhibitory effect of antibodies against cytochrome P-450$_{TYR}$

The experiments are carried out using monospecific antibodies against P-450$_{TYR}$ as described in section 6.4. which uses antibodies against cytochrome P-450$_{Ox}$. The results are similiar to those obtained with the antibody against cytochrome P-450$_{Ox}$ except that the cytochrome P-450$_{TYR}$ antibody exerts a stronger inhibitory effect (up to 60%) on cyanide production.

5.5. cDNA libraries and colony screening

Poly A$^+$ RNA is isolated from 3 cm high etiolated seedlings of Sorghum bicolor (L.) Moench grown as described for seedlings used for preparation of microsomes. The poly A$^+$ RNA is used for the construction of a λgt11 expression library and a λgt10 library. The construction of the libraries can be done according to the procedures described for example in Sambrook et al, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989 or can be ordered by commercial companies such as Strategene (La Jolla, Calif.). Antibodies obtained against cytochrome P450TYR are used to screen the expression libraries as described by Young et al, Procl. Natl. Acad. Sci USA 82:2583–2587. Antigen-antibody complexes are detected enzymatically with alkaline phosphatase-conjugated antibodies (Dakopatts). DNA from 4 positive plaques is prepared according to Grossberger, Nucleic Acid Research 15:6737, 1987. Inserts from λ phages are subcloned into pBluescript II SK (Strategene). Comparison of the deduced amino acid sequence from one of the four inserts with the amino acid sequences obtained from protein sequencing of P450$_{TYR}$ shows that this clone is a partial cDNA clone for P450$_{TYR}$. The partial cDNA clone is used as a probe for a new screen of the λgt10 and λgt11 libraries. The insert sizes of 45 positive clones are determined by southern blotting. Additionally the 45 positive clones are examined for hybridization with two different mixtures of oligonucleotides by southern blotting. The sequences of the oligonucleotide mixtures are based on the partial amino acid sequence data and specify a sequence near the N-terminal end (amino acids 4 to 9) and a sequence near the C-terminal end (amino acids 533–538). Oligonucleotide synthesis is carried out on a Cyclone Plus DNA Synthesizer. Sequencing of one clone derived from the λgt10 library showing the expected size and hybridizing with the two oligonucleotide mixtures shows that the clone is a full-length cDNA clone encoding cytochrome P450$_{TYR}$.

Oligonucleotide specifying amino acids 4 to 9 (MEVEAA) (SEQ ID NO: 17) 5'-ATG-GA[G,A]-GT[C,G,T,A]-GA[G,A]-GC[CGTA]-GC-3'

Oligonucleotide specifying amino acids 533 to 538 (DFTMAT) (SEQ ID NO: 18) 5'-GA[C,T]-AC[C,G,T,A]-TT[C,T]-ATG-GC[C,G,T,A]-AC-3'

5.6. DNA sequencing

DNA sequencing is carried out by the dideoxy chain method (Sanger et al, Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977) using [^{35}S]-dATP. T7 DNA polymerase and deoxynucleotides are obtained from Pharmacia, dideoxynucleotides from Boehringer Mannheim. Sequencing of the full-length cDNA clone is done partly by subcloning and partly by using synthetic oligonuecleotides as primers. The oligonucleotide primers can be ordered with commerical companies.

5.7. Southern blotting

λDNA isolated from the positive clones is digested with Eco R1. The inserts are separated from λDNA by electrophoresis on a 0.7% agarose gel. After electrophoresis, DNA is capillary blotted onto a Zetaprobe membrane (Biorad) using 10 mM NaOH for the transfer. Hybridization is performed at 68° C. in 1.5× SSPE (270 mM NaCl, 15 mM Na_2HPO_4 pH 7.0, 1.5 mM EDTA, 1% sodium dodecyl sulphate) 10% dextransulphate, 0.5% skim milk and 0.1 mg/ml salmon sperm DNA for 16 hours. When the partial cDNA clone is used as probe for hybridization it is labeled with [α-^{32}P]dCIP using a random prime labelling kit (Amersham International plc.). The oligonucleotide mixtures are 5'end labeled according to Okkels et al. (Okkels et al, FEBS Letters 237:108–112, 1988). The filters are washed first in 2× SSC (0.9M NaCl, 0.09M trisodium citrate, 0.1% SDS) (SEQ ID NO: 24) at 47° C. for 15 min., then in a fresh solution of the same composition at 56° C. for 15 min. and finally in 0.1× SSPE, 0.5% SDS for 30 minutes at 65° C. The presence of radioactively labelled hybridization bands on the filter is monitored by X-ray autoradiography.

5.8. Characterization of a full-length cDNA clone

λDNA isolated from the positive clones is digested with Eco RI. The insert is separated from λDNA by electrophoresis on a 0.7% agarose gel and subcloned into the EcoR1 site of the vector pBluescript SK (Strategene) contained within the sequence GCAGGAATTCCGG. The four last bases of this sequence are listed as the first four bases in SEQ ID NO: 1. A clone comprising the described cDNA has been deposited with the Agricultural Research Culture Collection (NRRL), 1815 N. University Street Peoria, Ill. 61604 U.S.A. under the accession number NRRL B-21168.

The orientation of the insert in the vector is determined as the polylinker restriction site for Pst I being adjacent to the 5'end of the cytochrome P-450 sequence. The sequence of the insert is shown in SEQ ID NO: 1. The sequence comprises an open reading frame (ORF) starting at nucleotide 188 and ending at nucleotide 1861 of SEQ ID NO: 1. It encodes a protein of 558 amino acids and a molecular mass of 61887 Da shown in SEQ ID NO: 2. The sequence comprises the sequences of SEQ ID NO: 3 to SEQ ID NO: 11. The protein is not subject to post-translational modification at the N- and C-terminal ends except for the removal of the N-terminal methionine residue. The N-termninal region of cytochrome P-450$_{TYR}$, however, shows four motifs which in animals are known to target heme-thiolate proteins to the endoplasmatic reticulum.

Searches for sequence similarity are made using the programmes BLAST and FASTA in the nucleotide sequence data bases provided by the EMBL. Pairwise comparisons of cytochrome P-450$_{TYR}$ with other cytochrome P-450 sequences are performed using the programme GAP of the Genetics Computer Group GCG sofware package. Multiple alignments were made using the GCG programme PILEUP.

5.9. Expression of native cytochrome P-450$_{TYR}$ in *E. coli*

Plasmid pCWOri+ (Gegner et al, Prod. Natl. Acad. Sci. USA 88:750–754, 1991) is used to express the wildtype cytochrome P-450$_{TYR}$ cDNA sequence as described by Barnes et al, Prod. Natl. Acad. Sci. USA 80:5597–5601, 1991. cDNA sequences are introduced into the expression plasmid using polymerase chain reaction (PCR) mutagenesis. A synthetic oligonucleodde (TYROL1b) containing an amino acid-conserving and nucleotide modifying 5' cDNA sequence is used in conjunction with a downstream oligonucleotide (TYROL3) to amplify the N-terminal sequence between the ATG initiator codon (contained within an NdeI site) and a unique BamHI restriction site within the cytochrome P-450$_{TYR}$ sequence. A synthetic oligonucleotide (TYROL2) is used in conjunction with an oligonucleotide (TYROL4) complementary to a unique PstI restriction site to introduce a HindIII restriction site immediately downstream of the TGA stop codon. The expression plasmid pCWtyr is constructed by simultaneous ligation of the 278 basepair PCR NdeI/BamHI fragment, the 1257 basepair BamHI/PstI fragment of cytochrome P-450$_{TYR}$ and the 146 basepair PCR PstI/HindIII fragment with the NdeI/HindIII cleaved vector DNA. *E. coli* strain JM 109 transformed with plasmid pCWtyr is grown in LB/ampicilline medium at 37° C. Expression of cytochrome P-450$_{TYR}$ is obtained by growing the cells in a medium containing 1 mM isopropyl beta-D-thiogalactopyranoside (IPTG) and shifting the cells to growth at 28° C. at 125 rpm. *E. coli* produces a functionally active cytochrome P-450$_{TYR}$ enzyme which converts tyrosine into oxime. The analytical procedures are as in the reconstitution experiments described in section 5.3. above. The expressed cDNA clone encoding P-450$_{TYR}$ specifies the synthesis of a single cytochrome P-450 enzyme. Since this enzyme catalyzes the conversion of tyrosine all the way to p-hydroxyphenylacetaldehyde oxime, this unambigeously demonstrates that cytochrome P-450$_{TYR}$ is multifunctional.

The following oligonucleotides are used:

TYROL1b (SEQ ID NO: 19)

5'-CGG GAT CCA TAT GCT GCT GTT ATT AGC AGT TTT TCT GTC GTA-3'

TYROL2 (SEQ ID NO: 20)

5'-GAC CGG CCG AAG CTT TAA TTA GAT GGA GAT GGA-3'

TYROL3 (SEQ ID NO: 21)

5'-AGT GGA TCC AGC GGA ATG CCG GCT T-3'

TYROL4 (SEQ ID NO: 22)

5'-CGT CAT GCT CTT CGG AA-3'

5.10. Expression of truncated and modified cytochrome P-450$_{TYR}$ in *E. coli*

A modified cytocrome P-450$_{TYR}$, in which the 35 N-terminal amino acids are replaced by the nine N-terminal amino acids from bovine 17α hydroxylase is introduced into the expression vector pSP19g10L which can be obtained from Dr. Henry Barnes (La Jolla, Calif.). This plasmid contains the lac Z promoter fused with the known short leader sequence (g10L) of gene 10 from bacteriophage T_7 (Olin et al, 1988). A construct containing the N-terminal amino acids from bovine 17α hydroxylase and a truncated form of the P-450$_{TYR}$ gene is designed using PCR mutagenesis: Oligonucleotide TYROL 1d (5'-CGG GAT CCA TAT GGC TCT GTT ATT AGC AGT TTT TCT GTC GTA CCT GGC CCG-3'; SEQ ID NO 23) containing a 5' mutant cDNA sequence as well as a BamHI and NdeI restriction site is used together with oligonucleotide TYROL 3 comprising the sequence surrounding the unique BamHI restriction site downstream of the ATG start codon of the cDNA coding for P-450$_{TYR}$ to amplify a modified N-terminal sequence of P-450$_{TYR}$. The amplification product is cut with NdeI and BamHI restriction enzymes. To introduce a HindIII site immediately downstream of the stop codon of the P-450$_{TYR}$ gene, oligonucleotides TYROL 2 and TYROL 4 are used in a polymerase chain reaction to obtain a C-terminal fragment of the P-450$_{TYR}$ gene comprising a HindIII restriction site immediately downstream of the stop codon. The amplification product is cut with PstI and HindIII restriction enzymes. The complete expression plasmid is constructed by simultaneous ligation of the N-terminal NdeI/BamHI fragment, the BamHI/PstI fragment of the P-450$_{TYR}$ gene and the C-terminal PstI/HindIII fragment into NdeI/HindIII cleaved pSP19g10L vector DNA. The expression vector obtained is transformed into *E. coli* strain JM109. Transformed *E. coli* produce 300 nmol to 500 nmol cytochrome P-450$_{TYR}$ per liter cell culture upon growth at 28° C. in the presence of 1 mM isopropyl-p-D-thiogalactopyranoside and at 125 rpm. Expression levels as high as 900 nmol per liter, equivalent to 55 mg P-450$_{TYR}$ per liter, have been obtained.

Administration of tyrosine to the cell culture results in the production of p-hydroxyphenylacetaldehyde oxime, whereas a cell culture transformed with pSP19g10L alone does not produce the oxime.

Reconstitution experiments with E. cole-expressed cytochrome P-450$_{TYR}$ and sorghum NADPH cytochorme P450 reductase in dilaurylpohsphatidylcholine micelles is performed as described in section 5.3. above. Turnover rates of 349 nmol oxime per nmol P-450$_{TYR}$ per minute can be demonstrated, which is equivalent to the values obtained with sorghum P-450$_{TYR}$.

Purified sorghum cytochrome P-450$_{TYR}$ can be shown to form type I substrate binding spectra with tyrosine and N-hydroxytyrosine (compare section 5.1.). Using P-450$_{TYR}$

expressed in *E. coli* it can be shown that in addition to tyrosine and N-hydroxytyrosine P-450$_{TYR}$ is also able to form a type Ispectrum with p-hydroxyphenylacetaldehyde oxime, 2-nitro-(p-hydroxyphenyl)ethane, p-hydroxyphenylacetonitrile as well as phenylalanine. The molar extinction coefficient $E_{420-390}$ for tyrosine and N-hydroxytyrosine as genuine substrates of P-450$_{TYR}$ are 75.8 $cm^{-1}mM^{-1}$ and 64.6 $cm^{-1}mM^{-1}$, respectively, whereas the extinction coefficients of the other compounds vary from 20–40 $cm^{-1}mM^{-1}$.

Reconstitution experiments using phenylalanine as substrate do not result in the production of the corresponding oxime. This indicates, that cytochrome P-450$_{TYR}$ has a narrow substrate specificity with respect to its enzymatic activity although it is able to bind many tyrosine analogues.

Administration of ^{14}C-tyrosine directly to *E. coli* cells expressing cytochrome P-450$_{TYR}$ results in the production of p-hydroxyphenylacetaldehyde oxime, indicating that *E. coli* is able to provide the reducing equivalents for cytochrome P-450$_{TYR}$.

The following oligonucleotides are used:

TYROL1d (SEQ ID NO: 23)
5'-CGG GAT CCA TAT GGC TCT GTT ATT AGC AGT TT TCT GTC GTA CCT GGC CCG-3'
TYROL2 (SEQ ID NO: 20)
5'-GAC CGG CCG AAG CTT TAA TTA GAT GGA GAT GGA-3'
TYROL3 (SEQ ID NO: 21)
5'-AGT GGA TCC AGC GGA ATG CCG GCT T-3'
TYROL4 (SEQ ID NO: 22)
5'-CGT CAT GCT CTT CGG AA-3'

Example 6

Characterization of P-450$_{Ox}$

6.1. Substrate binding spectra of P450$_{Ox}$

Similiar experiments as reported in section 5.1 are carried out using isolated cytochrome P-450$_{Ox}$ with p-hydroxyphenylacetaldoxime and p-hydroxyphenylacetonitrile as substrate. Cytochrome P-450$_{Ox}$ is found to be multifunctional as P-450$_{TYR}$. Isolated cytochrome P-450$_{Ox}$ resembles the cytochrome P-450 reported to convert oximes to nitriles in rat liver microsomes (DeMaster et al, J. Org. Chem. 5074–5075, 1992).

6.2. Molecular weight and Amino acid sequence data

The molecular weight of P-450$_{Ox}$ as determined by SDS-PAGE is 51 kD. Amino acid sequences are obtained by automated Edman degradation. The internal polypeptides are obtained by trypsin digestion of the purified protein and subsequent separation of peptides using reverse phase HPLC.

N-terminal sequence:
--MDLADIPKQQRLMAGNALVV-- (SEQ ID NO: 12)
Additional peptide sequences:
--ARLAEIFATII-- (SEQ ID NO: 13)
--EDFTVTTK-- (SEQ ID NO: 14)
--QYAALGSVFTVPII-- (SEQ ID NO: 15)
--XXPFPI-- (SEQ ID NO: 16)

6.3. Reconstitution of cytochrome P-450$_{Ox}$ activity:

The reconstitution assay used for cytochrome P-450$_{Ox}$ is similiar to that used for cytochrome P-450$_{TYR}$ described in section 5.3. A typical assay contains 10 μl DLPC (10 mg/ml); 50 μl cytochrome P-450$_{Ox}$ (24 μg/ml); 50 μl NADPH-cytochrome P-450 oxidoreductase; 20 μl of either 10 mM p-hydroxyohenylacetonitrile or p-hydroxyphenylacetaldehyde oxime; 10 μl NADPH (25 mg/ml); and 60 μl potassium phosphate buffer (pH 7.9).

The reconstitution assay demonstrates that cytochrome P-450$_{Ox}$ converts p-hydroxyphenylacetaldehyde oxime to p-hydroxymandelonitrile. The analytical procedures used are those described for cytochrome P-450$_{TYR}$.

6.4. Inhibitory effect of antibodies against cytochrome P-450$_{Ox}$

The effect of antibodies raised against cytochrome P-450$_{Ox}$ on the biosynthetic activity is measured as the decrease in cyanide production upon incubation of the sorghum microsomes with p-hydroxyphenylacetaldehyde oxime and p-hydroxybenzylcyanide as substrates. The composition of the 150 μl total volume reaction mixtures is: microsomes containing 33 μg protein, 1.5 μmol substrate, 7.5 μmol tricine pH 8.0, 0.33 μmol NADPH, 0–255 μg antibodies and 0–255 μg reference immunoglobulin. The total amount of immunoglobulin in the assay is in each sample adjusted to 225 μg using purified immunoglobulin from a nonimmunized rabbit. The antibodies are preincubated with the microsomes for 15 minute at 30° C. before substrate and NADPH are added. Subsequently the reaction is incubated at 30° C. for 30 minutes. Cyanide is determined by the König reaction (König, Z. Angew. Chem. 18:115, 1905) using methodology described in Halkier and Moller, Plant Physiol. 90:1552–1559, 1989. A value of $A_{680-585}$=1.5 corresponds to 10 nmoles cyanide. Protein concentration was determined using the method of Bradford (Bradford, Anal. Biochem. 72:248–254, 1976). A typical result of such an inhibition experiment is shown in Table E.

TABLE E

μg antibody	0	15	30	60	120	225
Substrate p-hydroxyphenylacetaldehyde oxime						
$A_{680-585}$	1,00	0.96	0,92	0.87	0,88	0,66
inhibition	0%	4%	8%	13%	12%	34%
Substrate p-hydroxybenzylcyanide						
$A_{680-585}$	1,07	1,03	1,03	0,88	0,87	0,70
inhibition	0%	4%	4%	18%	18%	35%

The data show, that the antibody inhibits the reactions to the same extent whichever substrate is added to the microsomal preparation.

Example 7

Induction of glucosinolate production in *Tropaeolum majus*.

Seeds of *Tropaeolum majus L.* cv Empress of India (Dansk Havefroforsyning, Kolding, DK) are allowed to imbibe and germinate in complete darkness for one week at 25° C. In vivo biosynthesis experiments are performed wherein 1 μCi of the tracer ^{14}C-labelled phenylalanine is administered to excised dark-grown seedlings for 24 hours followed by boiling of the plant material in 90% methanol and analysis of the extracts by HPLC as described by Lykkesfeldt and Moller, 1993. Prior to administration of the tracer to the excised seedling or leaf, the intact plant is subjected to a potential inducer for 24 hours. Administration of 10 mM phenylalanine or 2% ethanol to the vermiculate in which the etiolated seedlings are grown results in a threefold increase in glucosinolate production as compared to control experiments with water. Spraying with 100 μM jasmonic acid followed by incubation for 24 hours results in a fivefold induction in etiolated seedlings and green leaves.

Example 8

Preparation of biosynthetically active microsomes from glucosinolate-producing plant meterial

The biosynthetic pathways of glucosinolates, and cyanogenic glucosides share homology by having amino acids as

precursors and oximes as intermediates. The assignment of amino acids and oximes as precursors and intermediates in the glucosinolate biosynthetic pathway is based on in vivo experiments demonstrating that these compounds are efficient precursors for glucosinolates. In vitro biosynthetic studies have hitherto not been possible due to the detrimental effect of the degradation products of glucosinolates on enzyme activities. The degradation products are formed upon disruption of the cellular structure. In the disrupted tissue, the glucosinolate-degrading enzyme myrosinase gets in contact with the glucosinolates resulting in the generation of isothiocyanates inactivating the enzymes. We demonstrate that microsomal preparations isolated form one week old plants of either *Sinapis alba* or *Trapaeolum majus* are able to convert tyrosine and phenylalanine, respectively, to the corresponding oximes. The enzymatically active microsomal preparations are obtained by using an isolation buffer fortified with 100 mM ascorbic acid known to inhibit the activity of myrosinase and by inducing the glucosinolate-producing enzyme system prior to the preparation of microsomes. The glucosinolate-producing enzyme systems are induced by taking 7-days-old dark-grown Sinapis plants or 3–4 weeks old light-grown Tropaeolum plants and placing them in the light for 3 days. During this three day period, the young plants are sprayed with 50 μM jasmonic acid once a day. After 3 days of induction, the plants are harvested and microsomes are prepared as described in section 5.1, except that the homogenisation buffer consists of 250 mM Tricine pH 7.9, 250 mM sucrose, 50 mM sodium bisulfite, 100 mM ascrobic acid, 4 mM DDT, 2 mM EDTA, 1 mM PMSF, and 5 mg/ml BSA. The microsomal preparation is dialysed against homogenization buffer for 1 hour, followed by dialysis against 50 mM Tricine pH 7.9 and 2 mM DTT for another hour.

Example 9

In vitro biosynthesis of oxime by extracts from glucosinolate-containing plants

The microsomal reaction mixture consists of 80 μl microsomes (10 mg protein per ml), 10 μl ^{14}C-phenylalanine (0.5 μCi, 464 mCi/mmol, Amersham) or ^{14}C-tyrosine (0.5 μCi, 450 mCi/mmol, Amersham) and 10 μl NADPH (75 mg/ml). The reaction mixtures are incubated for 1 hour at 37° C. At the end of the incubation period, the reaction mixtures are extracted with 1500 μl ethyl acetate. The ethyl acetate phase is evaporated to dryness, redissolved in a small volume and analyzed. The production of oximes in the microsomal reaction mixtures can be demonstrated by thin layer chromatography as well as by HPLC analysis as described in section 5.3.

Example 10

Involvement of cytochrome P450-dependent monooxygenases in the glucosinolate pathway

Based on the similarity between the first part of the biosynthetic pathways of glucosinolates and cyanogenic glucosides, it was anticipated that the conversion of amino acid to oxime in the glucosinolate pathway is catalyzed by a multifunctional cytochrome P450 monooxygenase homologous to P450$_1$ in the cyanogenic glucoside pathway. In vivo experiments, where radioactively labelled phenylalanine is administered to etiolated tropaeolum seedlings in the presence and absence of 1 mM of the cytochrome P450 inhibitors enilketoconazol and tetcyclacis demonstrate that cytochrome P450 inhibitors cause a reduction of glucosinolate without causing a reduction in the uptake of phenylalanine as measured by ethanol extraction of the plant material. This indicates that the biosynthesis of glucosinolates is dependent on cytochrome P450.

Direct demonstration of the involvment of cytochrome P450 in glucosinolate biosynthesis can be obtained using the in vitro microsomal enzyme system from tropaeolum to demonstrate photoreversible carbon monoxide inhibition of oxime production. The microsomal reaction mixtures are incubated using different experimental conditions. The reaction mixtures are analyzed by HPLC.

Experimental condition	% inhibition of oxime production
O_2 without light	0
O_2 with light	11
CO/O_2 without light	65
CO/O_2 with light	23

The possibility to reactivate the microsomal enzyme system upon irradiation with 450 nm light shows, that the conversion of phenylalanine to the corresponding oxime in the biosynthetic pathway of glucosinolate is dependent on cytochrome P450.

Example 11

Toxicity of cyanogenic glycosides for insects.

Insects or insect larvae are fed on a diet containing added cyanogenic glycoside, a diet containing added cyanogenic glycoside and callus, or a diet supplemented with the supernatant of callus ground-up in the presence of the cyanogenic glycoside. Mortality is compared to the mortality of insects or insect larvae fed on the diet only.

Example 12

Activity of Amygdalin on larval mortality of Western Corn Root Worm (WCRW):

WCRW larvae are fed on a diet with added amygdalin, a diet with added amygdalin and Black Mexican Sweet (BMS) callus or on a diet supplemented with the supernatant of BMS-callus ground-up in the presence of amygdalin. Larval mortality is compared to the mortality of larvae fed on the diet only.

The results show that amygdalin is lethal in the presence of BMS-callus with an LC_{50} of ~1 mg/ml, and that it is lethal at 2 mg/ml in the absence of BMS-callus. There is significantly less lethality at amygdalin concentrations of less than 1 mg/ml when BMS-callus is absent.

Example 13

Activity of Dhurrin on larval mortality of Western Corn Root Worm (WCRW):

The activity of dhurrin on larval mortality of WCRW was determined as described for amygdalin in example 12.

The results show that the LC_{50} of dhurrin is 368 μg/ml with 95% confidence limits of 0.28–0.48 μg/ml. The slope of the regression line is 2.5.

Example 14

Transfection of maize by direct Bombarding of Immature Zygotic Embryos and Isolation of Transformed Callus with the Use of Phosphinothricin as a selection agent.

Immature embryos are obtained approximately 14 days after self-pollination. The immature zygotic embryos are divided among different target plates containing medium capable of inducing and supporting embryogenic callus formation at 36 immature embryos per plate. The immature zygotic embryos are bombarded with plasmids encoding a cytochrome P-450 monooxygenase and a chimeric gene coding for resistance to phosphinothricin using the PDS-1000/He device from DuPont The plasmids are precipitated

onto 1 μm gold particles essentially according DuPont's procedure. Each target plate is shot one time with the plasmid and gold preparation and phosphinothricin is used to select transformed cells in vitro. Selection is applied at 3 mg/i one day after bombardment and maintained for a total of 12 weeks. The embryogenic callus so obtained is regenerated in the absence of the selection agent phosphinothricin. The regenerated plants are tested for their resistance to insects, acarids or nematodes.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

TABLE A

PLANT CLASSIFICATION ACCORDING TO USE

CEREALS

Monocot

Avena nuda (chinensis)	Chines naked oat
A. sativa	Common oats
Eleusine coracan	African millet
Eragrostis tef	Tef grass
Fagopyrum esculentum	Buckwheat
F. tataricum	Rye buckwheat
Hordeum distichum	Two-row barley
H. vulgare	Barley
Oryza sativa	Rice
Panicum italicium	Italian millet
P. miliaceum	Broomcorn millet
Pennisetum glaucum	Spiked millet
P. spicatum (americanum)	Perl millet
Secale cereale	Rye
Sorghum vulgare	Grain sorghums
X Triticosecale	Triticale
Triticum aestivum	Common wheat
T. dicoccum	Emmer
T. durum	Abyssinian hard wheat
T. monococcum	Einkorn wheat
Zea mays	Corn, sweet corn

Dicot

Amaranthus paniculatus	Rispenfuchsschwanz
Fagopyrum esculentum	Buchweizen
F. tataricum	

PROTEIN CROPS

Dicot

Arachis hypogea	Groundnut, peanut
Cajanus indicus	Pigeon pea
Cicer arietinum	Chickpea
Dolichos lablab	Hyacinth bean
Glycine gracilis	Manchurian Soya
G. max	Soyabean
G. ussuriensis	Wild soya
Lathyrus sativus	Grass pea
Lens culinaris	Lentil
Mucuna pruriens	Cowitch, Florida velvet bean
Phaseolus acutifolius	Tepary bean
P. aureus	Mung, green gram
P. lunatus	Lima bean, Sieva
P. coccineus (multiflorus)	Scarlet runner bean
P. mungo	Black gram
P. vulgaris	French, common, kidney or dwarf bean
Vicia faba	Horse bean, broad bean
Vigna angularis	Adzuki bean
V. sesquipedalis	Asparagus (yard-long bean)
V. sinensis	Cowpea

TABLE A-continued

PLANT CLASSIFICATION ACCORDING TO USE

FRUIT CROPS

Dicot

Amygdalus communis	Almond
Ananas comosus	Pineapple
Artocarpus communis	Breadfruit
Carica papaya	Papaya
Citrullus vulgaris	Watermelon
Citrus grandis	Pummelo
C. medica	Citron, lemon
C. nobilis	Tangerine
C. reticulata	Mandarin
C. sinensis	Orange
Cydonia oblonga	Quince
Diospyros kaki	Japanese persimmon
Ficus carica	Fig
Fragaria chiloensis	Wild strawberry
F. virginiana	Strawberry
Litchi chinensis	Litchi
Malus asiatica	Chines apple
M. pumila	Appple
Mangifera indica	Mango
Morus rubra	Red mulberry
Musa cavendishii	Banana
M. paradisiaca	Banana
Passiflora edulis	Passion fruit, purple granadilla
P. ligularis	Passion flower
Persea americana	Avocado pear
Phoenix dactylifera	Date palm
Prunus armeniaca	Apricot
P. avium	Sweet cherry, mazzard
P. cerasifera (divaricata)	Cherry plum
P. cerasus	Cherry
P. domestica	European plum or prune
P. maheleb	Maheleb cherry
P. persica	Peach and nectarine
P. pseudocerasus	Cherry
P. salicinia	Japanese peach
P. serotina	Wild black cherry
Psidium guajava	Guava
Punica granatum	Pomegranate
Pyrus communis	Pear
P. ussuriensis	Chinese pear
Ribes grossularia	Gooseberry
R. nigrum	Black currant
R. rubrum	Red and white currant
Rubus idaeus	European raspberry
R. strigosus	American raspberry
Tamarindus indica	Tamarind
Vaccinium angustifolium	Sugarberry
V. ashei	Rabbiteye blueberry
V. corymbosum	Highbush blueberry
V. myrtilloides	Canada blueberry
V. oxycoccos	Cranberry
Viburnum trilobum	American cranberry bush
Vitris labrusca	Fox grape
V. vinifera	Grape

VEGETABLES AND TUBERS

Monocot

Allium ascalonicum	Shallot, breen onion
A. cepa	Onion
A. chinense	Onion
A. fistulosum	Welch onion
A. porrum	Leek
A. sativum	Garlic
A. schoenoprasum	Chives
Asparagus officinalis	Asparagus (var. *attilis*)
Zea mays	sweet corn

Dicot

Amoracia lapathifolia	Horseradish
Apium graveolens	Celery

TABLE A-continued

PLANT CLASSIFICATION ACCORDING TO USE

Arabidopsis thaliana	Common wall cress
Beta vulgaris	Sugar, mangold or garden beet
Brassica alboglabra	Chinese kale
B. campestris	Turnip rape
B. carinata	Ambyssian mustard
B. cernea	Karashina
B. chinensis	Chinese mustard or pak-choi
B. hirta	White mustard
B. juncea	Pai, brown mustard, Indian mustard
B. kaber	Charlock
B. napobrassica	Swede or rutabaga
B. napus	Rape, oil rape, kale
B. nigra	Black mustard
B. oleracea	Cole, kale, collards, brussels sprouts, cauliflower, cabbage, kohlrabi, broccoli
B. pekinensis	Chines cabbage or celery cabbage
B. rapa	Turnip
Cajanus cajan (indicus)	Pigeon pea
Canavalia ensiformis	Jack bean
Canna edulis	Edible canna
Capsicum annuum	Common cultivated pepper
C. chinense	Pepper
C. frutescens	Cayenne pepper
C. pendulum	Pepper
C. pubescens	Pepper
Cichorium endivia	Endive
C. intybus	Chicory
Colocasia antiquorum	Taro
Crambe maritima	Sea kale
Cucumis melo	Melon, cantaloupe
C. sativus	Cucumber
Cucurbita ficifolia	Malabar gourd
C. foetidissima	Calabazilla, buffalo gourd
C. maxima	Pumpkin
C. moschata	Winter pumpkin
C. pepo	Summer squash, vegetable marrow
Cynara scolymus	Globe artochoke
Daucus carota	Carrot
Dioscorea alata	Yam
D. batatas	Chines yam
D. cavennensis	Attoto yam
Eruca sativa Mill.	Rocket salad, rocket or roquette
Ipomea batatas	Sweet potato
Lactuca sativa	Lettuce
Lepidium sativum	Garden cress
Lycopersicon cerasiforme	Cherry tomato
L. esculentum	Tomato
Manihot esculenta	Manioc, cassava
Nasturtium officinale	Water cress
Pastinaca sative	Parsnip
Petroselinum crispum (sativum)	Parsley
Physalis peruviana	Ground cherry
Pisum sativum	Pea
Raphanus sativus	Radish
Rheum officinale	Rhubarb
R. rhaponticum	English rhabarb
Scorzonera hispanica	Black salsify
Sechium edule	Chayote
Solanum andigenum	Andean potato
S. melongena	Eggplant
S. muricatum	Pepino
S. phureja	Potato
S. tuberosum	Common potato
Psinacia oleracea	Spinach

TABLE A-continued

PLANT CLASSIFICATION ACCORDING TO USE

NUTS

Dicot

Anacardium occidentale	Cashew
Arachis hypogaea	Peanut
Carya illinoinensis	Pecan
C. ovata	Shagbark hickory
Castanea sativa	Chestnut
Covos nucifera	coconut palm
Corylus americana	American hazel, filbert
C. avellana	European hazel, cobnut
Juglans nigra	Black walnut
J. regia	English walnut
J. sinensis	Walnut
Litchi chinensis	Litchi
Macadamia integrifolia	Queensland nut
Pistacia vera	Pistachio nut
Prunus amygdalus	Almond

OIL CROPS

Monocot

Zea mays	Corn

Dicot

Aleurites cordata	Tung, China wood oil
A. moluccana (triloba)	Candlenut
Arachis hypogea	Ground nut, penut
brassica campestris	Rapeseed oil, canola oil
B. napus	Rapeseed oil, canona oil
Cannabis sativa	Hampseed oil
Carthamus tinctorius	Safflower oil
Cocos nucifera	Coconut palm
Elaeis guineensis	Oil palm
Glycine gracilus	Manch, soya
G. max	Soybean
G. ussuriensis	Wild soya
Cossypium hirsutum	Cottonseed oil
Helianthus annus	Sunflower
Linum usitatissimum	Flax
Olea europaea	Olive
Papaver somniferum	Poppy seed
Ricinus communis	Castor bean
Sesamum indicum	Sesame

SUGAR CROPS

Monocot

Saccharum officinarum (*officinarum* x *spontaneum*)	Sugarcane
S. robustum	
S. sinense	Sugarcane
S. spontaneum	Kans grass
Sorghum dochna	Sorgo syrup, sugar sorghum

Dicot

Acer saccharum	Sugar maple
Beta vulgaris	Sugar or mangold beet

FORAGE AND TURF GRASSES

Monocot

Agropyron cristatum	Crested wheatgrass
A. desertorum	Crested wheatgrass
A. elongatum	Tall wheatgrass
A. intermedium	Intermediate wheatgrass
A. smithii	Western wheatgrass
A. spicatum	Blue bunch wheatgrass
A. trachycaulum	Slender wheatgrass
A. trichophorum	Pubescen wheatgrass
Alopecurus pratensis	Meadow foxtail
Andropogon gerardi	Big bluestem
Arrhenatherum elatius	Tall oat grass
Bothrichloa barbinodis	Cane blestem
B. ischaemum	King ranch bluestem

TABLE A-continued

PLANT CLASSIFICATION ACCORDING TO USE

B. saccharoides	Silver bluestem
Bouteloua curipendula	Side oats grama
B. eriopoda	Black grama
B. gracilis	Blue grama
Bromus erectus	Upright brome
B. inermis	Smooth brome
B. riparius	Meadow brome
Cenchrus ciliaris	Buffel grass
Chloris gayana	Rhodes grass
Cymbopogon nardus	Citronella grass
Cynodon dactylon	Bermuda grass
Dactylis glomerata	Cocksfoot
Dichanthium annulatum	Kleberg bluestem
D. aristatum	Angleton bluestem
D. sericeum	Silky bluestem
Digitaria decumbens	Pangola grass
D. smutsii	
Elymus angustus	Altai wild rye
E. junceus	Russian wild rye
Eragrostis curvula	Weeping love grass
Festuca arundinacea	Tall fescue
F. ovina	Sheeps fescue
F. pratensis	Meadow fescue
F. rubra	Red fescue
Lolium multiflorum	Italian ryegrass
L. perenne	Perennial ryegrass
Panicum maximum	Guinea grass
P. purpurascens	Para grass
P. virgatum	Switchgrass
Paspalum dilatatum	Dallis grass, large water grass
P. notatum	Bahia grass
Pennisetum clandestinum	Kikuyu grass
P. purpureum	Dry napier grass
Phalaris arundinacea	Reed canary grass
Phleum bertolinii	Timothy
P. pratense	Timothy
Poa fendleriana	Mutton grass
P. nemoralis	Wood meadow grass
P. pratensis	Kentucky bluegrass
Setaria sphacelata	Rhodesian timothy
Sorghastrum nutans	Indian grass
Sorghum halepense	Johnson grass
S. sudanense	Sudan grass
Sorghum vulgare	Great millet
FORAGE LEGUMES	
Dicot	
Coronilla varia	Crown vetch
Crotalaria juncea	Sun hemp
Lespedeza stipulacea	Korean lespedeza
L. striata	Common lespedeza
L. sericea	
Lotus corniculatus	Birdsfoot trefoil
L. uliginosus	
Lupinus albus	Wolf bean, white lupin
L. angustifolius	Blue lupin
L. luteus	European yellow lupin
L. mutabilis	South American lupin
Medicago arabica	Spotted burr-clover
M. arborea	Tree alfalfa
M. falcata	Yellow lucerne
M. hispida	Caiifornia burr-clover
M. sativa	Alfalfa
M. tribuloides	Barrel medic
Melilotus albus	White sweet clover
M. officinalis	Yellow sweet clover
Onobrychis viciifolia	Sainfoin
Ornithopus sativus	Serradella
Pueraria thunbergiana	Kudzu vine
Trifolium alexandrinum	Egyptian clover
T. augustifolium	Fineleaf clover
T. diffusum	Rose clover
T. hybridum	Alsike clover
T. incarnatum	Crimson clover
T. ingrescens	Ball clover

TABLE A-continued

PLANT CLASSIFICATION ACCORDING TO USE

T. pratense	Red clover
T. repens	White clover
T. resupinatum	Persian clover
T. subterraneum	Subterranean clover
Trigonella foenumgraecum	Fenugreek
Vicia sative	Common vetch
V. villosa	Hairy vetch
V. atropurpurea	Purple vetch
V. angustifolia	Narrowleaf vetch
V. dasycarpa	Wooly pod vetch
V. ervilia	Monantha (bitter) vetch
V. pannonica	Hungarian vetch
V. calcarata	Bard vetch
FIBER PLANTS AND WOODY PLANTS	
Monocot	
Bambusa vulgaris	Bamboo
Dicot	
Agave sisalana	Sisal hemp
Boehmeria nivea	Rhea fiber, ramie
Cannabis indica	Hemp
C. sativa	Hemp
Ceiba pentandra	Silk cotton tree, kapok tree
Corchorus mucronata (striata)	Hemp
Gossypium arboreum	Tree cotton
G. barbadense	Egyptian cotton
G. herbaceum	Cotton
G. hirsutum	Upland cotton
G. [001b]nanking	Oriental cotton
Linum angustifolium	Wild flax
L. usitatissimum	Flax
Musa textiles	Manila hemp, abaca
DRUG CROPS	
Dicot	
Angelica archangelica	Angelica
Chrysanthemum cinerariifolium	Palm pyrethrum
Camellia sinensis	Chinese tea
C. coccineum	Pyrethrum
Coffea arabica	Coffee
C. canephora	Quillow coffee
Cola acuminata	Kola nut
Nicotiana rustica	Tobacco
N. tabacum	Tobacco
Papaver dubium	Poppy
P. somniferum	Opium poppy
Theobroma cacao	cocoa
SPICES AND FLAVORINGS	
Monocot	
Vanilla fragrans	Vanilla
Dicot	
Artemisa dracunculus	Tarragon
Cinnamomum zeylanicum	Cinnamon tree
Hibiscus esculentus	Okra
Salvia officinalis	Sage
Thymus vulgaris	Thyme
Pimpinella anisum	Anise
Mentha arvensis	Menthol
M. piperita	Peppermint
M. viridis	Spearmint
Coriandrum sativum	Coriander

Table B: REPRESENTATIVE PLANT PESTS

Coleoptera:

Diabrotica, Melanotus, Agriotes, Limonius, Dalopius, Eleodes, Chaetocnema, Macrodactylus, Sphenophorus, Sitophilus, Lisorhoptrus, Oulema, Rhyzopertha,

Prostephanus, Phyllophage, Cyclocephala, Popillia, Anthonomus, Zabrotes, Leptinotarsa
Lepidoptera:
Heliothis, Ostrinia, Diatraea, Elasmopalpus, Papaipema, Agrotis, Loxagrotis, Euxoa, Peridroma saucia, Chorizagrotis, Spodoptera, Pseudaletia, Chilo, Busseola, Sesamia, Eldana, Maliarpha, Scirpophaga, Duataea, Rupela, *Sitotroga cerealella,* Sitroga, *Plodia interpunctella,* Crambus, Mythimna, Nola, Pectinophora, Acontia, Trichoplusia, Anticarsia, Pseudoplusia, Manduca, Leptinotarsa, Lema
Thysanoptera:
Frankliniella, Anaphothrips, Hercothrips, Stenothrips
Homoptera:
Dalbulus, Cicadulina, Rhopalosiphum, Melanaphis, Anuraphis, Prosapia, Nilaparvata, Sogatella, Laodelphax, Sogatodes, Nephotettix, Reciian, Cofana, Empoasca, Poophilus, Schizaphis, Sipha, Paratrioza, Empoasca, Ophilia Scleroracus, Macrosteles, Circulifer, Aceratagallia, Agallia, Myzus, Macrosiphum, Aphis
Diptera:
Delia platura, Euxesta, Diopsis, Atherigona, Hydrellia, Orseolia, Chironomus, Contarinia
Orthoptera:
Melanoplus, Schistocerca, Sphenarium, Aneolamia
Isoptera:
Microtermes, Macrotemes, Allodontermes, Odontotermes
Heteroptera:
Nezara, Acrosternum, Euschistus, Blissus
Acarina:
Tetranychus, Paratetranychus, Oligonychus

LIST OF REFERENCES CITED

Barnes et al, Prod. Natl. Acad. Sci. USA 80:5597–5601, 1991
Bradford, Anal. Biochem. 72:248–254, 1976
Christou et, Plant Physiol. 87:671–674, 1988
Conn, Naturwissenschaften 66:28–34, 1979
Cronquist, 'The Evolution and Classification of Flowering Plants, New York Botanical Garden, Bronx, 1988
DeMaster et al, J. Org. Chem. 5074–5075, 1992
Food Chemical News 29:33.35, 1988
Gegner et al, Prod. Natl. Acad. Sci. USA 88:750–754, 1991
Grossberger, Nucleic Acid Research 15:6737, 1987
Halkier et al, 'Cyanogenic glucosides: the biosynthetic pathway and the enzyme system involve' in: 'Cyanide compounds in biology', Wiley Chichester (Ciba Foundation Symposium 140), pages 49–66, 1988
Halkier and Moller, Plant Physiol. 90:1552–1559, 1989
Halkier et al, The J. of Biol. Chem. 264:19487–19494, 1989
Halkier and Moller, Plant Physiol. 96:10–17, 1990
Halkier and Moller, The J. of Biol. Chem. 265:21114–21121, 1990
Halkier et al, Proc. Natl. Acad. Sci. USA 88:487–491, 1991
D'Halluin et al, The Plant Cell 4:1495–1505, 1992
Harlow and Lane, 'Antibodies A Laboratory Manual', Cold Spring Harbor Laboratory,
Harboe et al, 'A Manual of Quantitative Immunoelectrophoresis: Methods and Applications', Universitetsforlaget, Oslo, 1973
Horsch et al, Science 227:1229–1231, 1985
Ibenthal et al, Angew. Bot. 67:97–106, 1993
Jefcoate, Methods in Enzymology 52:258–279, 1978
Kindl et al, Phytochemistry 7:745–756, 1968
Koch et al, 8th Int. Conf. on Cytochrome P450, Abstract PII.053
Koch et al, Archives of Biochemistry and Biophysics, 292:141–150, 1992
König, Z. Angew. Chem. 18:115,1905
Koziel et al, Biotechnology 11: 194–200, 1993
Lieberei et al, Plant Phys. 90:3–36, 1989
Lüdtke et al, Biochem. Z. 324:433–442, 1953
Lykkesfeldt et al, Plant Physiology 102: 609–613, 1993
Matsuo et al, Phytochemistry 11:697–701, 1972
Murray et al, Nucleic Acids Research 17:477–498, 1989
Nebert et al, DNA Cell Biol. 10:1–14, 1991
Negrutiu et al, Plant Molecular Biology 8:363–373, 1987
Neuhaus et al, Theoretical and Applied Genetics 74:363–373, 1987
Okkels et al, FEBS Letters 237:108–112, 1988
Olin et al, Gene 73: 227–235, 1988
Omura et al, J. Biol. Chem. 239:2370–2378, 1964
Pourmohensi, PhD thesis, Gottingen, 1989
Sambrook et al, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989
Sanger et al, Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977
Schocher et al, Bio/Technology 4:1093–1096, 1986
Shillito et al, Biotechnology 3:1099–1103, 1985
Sibbesen et al, in: 'Biochemistry and Biophysics of cytochrome P-450. Structrue and Function, Biotechnological and Ecological Aspects', Archakov, A. I. (ed.), 1991
Sibbesen et al, 8th Int. Conf. on Cytochrome P450, Abstract PII.016
Underhill, Eur. J. Biochem. 2:61–63, 1967)
Underhill et al, Biochem. Soc. Symp. 38:303–326, 1973
Young et al, Procl. Natl. Acad. Sci USA 82:2583–2587
Wang et, Plant Molecular Biology 11:433–439, 1988

EP-240 208-A2	WO 89/05852	U.S. Pat. No. 5,023,179
EP-306,139-A	WO 89/07647	U.S. Pat. No. 5,231,020
EP 332 104	WO 91/13992	
EP-452 269-A2	WO 93/07278	
EP-458 367 A1	WO 92/09696	
	WO 90/11682	

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 24

(2) INFORMATION FOR SEQ ID NO: 1:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Sorghum bicolor

(vii) IMMEDIATE SOURCE:
(B) CLONE: P-450-Tyr

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCGGCTAGCT AGCTCATCGG GTGATCGATC AGTGAGCTCT CTCTTTGGCC TAGCTAGCTG 60

CTAGCAGTGC AGGTAGCCAA TCAAAGCAGA AGAACTCGAT CGATCGATCA TCACGATCGC 120

TGCTAGCTAG CTAGCTGCTC GCTCTCACAC TAGCTACGTG TTTTTGTTAA TTTGATATAT 180

ATATATAATG GCGACAATGG AGGTAGAGGC CGCGGCCGCC ACGGTGCTGG CCGCGCCCTT 240

GCTGTCCTCC TCCGCGATCC TCAAACTGCT GCTATTCGTA GTGACGCTCT CGTACCTGGC 300

CCGAGCCCTG AGGCGGCCAC GCAAAAGCAC CACCAAGTGC AGCAGCACAA CGTGCGCCTC 360

GCCCCCGGCC GGCGTTGGCA ACCCGCCGCT CCCACCGGGT CCCGTGCCGT GGCCCGTCGT 420

CGGCAACCTG CCGGAGATGC TGCTGAACAA GCCGGCATTC CGCTGGATCC ACCAGATGAT 480

GCGCGAGATG GGCACGGACA TCGCCTGCGT CAAGCTTGGC GGCGTCCACG TCGTGTCCAT 540

CACCTGCCCG GAGATCGCGC GGGAGGTGCT CCGGAAGCAG GACGCCAACT TCATATCCCG 600

CCCGCTCACC TTCGCCTCCG AGACGTTCAG CGGCGGGTAC CGGAACGCCG TGCTCTCGCC 660

CTACGGCGAC CAGTGGAAGA AGATGCGCCG CGTCCTCACC TCCGAGATCA TCTGCCCGTC 720

CCGCCACGCC TGGCTCCACG ACAAGCGCAC CGACGAGGCC GACAACCTCA CCCGCTACGT 780

CTACAACCTC GCCACCAAAG CCGCCACCGG CGACGTCGCC GTCGACGTCA GGCACGTCGC 840

TCGTCACTAT TGCGGCAACG TTATCCGCCG CCTCATGTTC AACAGGCGCT ACTTCGGCGA 900

GCCCCAGGCT GACGGCGGTC CGGGGCCGAT GGAGGTGCTG CATATGGACG CCGTGTTCAC 960

CTCCCTCGGC CTCCTCTACG CCTTCTGCGT CTCCGACTAC CTCCCCTGGC TGCGGGGCCT 1020

CGACCTCGAC GGCCACGAGA AGATCGTCAA GGAGGCTAAC GTGGCGGTGA ACAGGCTCCA 1080

CGACACGGTC ATCGACGACC GGTGGAGGCA GTGGAAGAGC GGCGAGCGGC AGGAGATGGA 1140

GGACTTCCTG GATGTGCTCA TCACTCTCAA GGACGCCCAG GGCAACCCGC TGCTGACCAT 1200

CGAGGAGGTC AAAGCGCAGT CACAGGACAT CACGTTCGCG GCGGTGGACA ACCCGTCGAA 1260

CGCCGTGGAG TGGGCGCTGG CAGAGATGGT GAACAACCCG GAGGTGATGG CGAAGGCGAT 1320

GGAGGAGCTG GACCGCGTCG TCGGACGGGA GAGGCTAGTG CAGGAGTCGG ACATTCCGAA 1380

GCTCAACTAC GTGAAGGCCT GCATCCGGGA GGCTTTCCGT CTGCACCCGG TGGCGCCCTT 1440

CAACGTGCCC CACGTCGCGC TCGCCGACAC CACCATCGCC GGCTACCGCG TTCCCAAGGG 1500

CAGCCACGTG ATCCTGAGCC GCACGGGGCT GGGCCGCAAC CCGCGCGTGT GGGACGAGCC 1560

CCTGCGCTTC TACCCGGACC GACACCTCGC CACCGCCGCG TCCGACGTCG CGCTCACCGA 1620

GAACGACCTG CGGTTCATCT CCTTCAGCAC CGGCCGCCGC GGCTGCATCG CCGCGTCGCT 1680

CGGCACCGCC ATGAGCGTCA TGCTCTTCGG AAGGCTCCTG CAGGGGTTCA CCTGGAGCAA 1740

-continued

GCCCGCCGGG GTGGAGGCCG TGGACCTCAG CGAGTCCAAG AGCGACACCT TCATGGCCAC 1800

CCCGCTGGTG CTGCACGCTG AGCCCAGGCT GCCGGCGCAC CTCTACCCGT CCATCTCCAT 1860

CTGATTAAAC GTACGGCCGG TCGTCATTAT ATTGTATGCA TATAATTAAA GACGAGCGAG 1920

CCTGCTGGTC ACACTTGCAT TGCATGTATC ATCAGCAGGG GGCTATGCAA TAAGTTTTTT 1980

TTTTCCGCGC TTGATTTCGT GGTGCTGTGC GTATTCTGCG CACACCGACT GTACGTACGA 2040

CGGCGTTCAG CTTTGTATTG TACCGAGTTA AAAAGTATTA TTATTATTAT CATCGACAAT 2100

AATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA 2143

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 558 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Sorghum bicolor

(vii) IMMEDIATE SOURCE:
(B) CLONE: P-450-Tyr

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Thr Met Glu Val Glu Ala Ala Ala Thr Val Leu Ala Ala
1 5 10 15

Pro Leu Leu Ser Ser Ser Ala Ile Leu Lys Leu Leu Leu Phe Val Val
20 25 30

Thr Leu Ser Tyr Leu Ala Arg Ala Leu Arg Arg Pro Arg Lys Ser Thr
35 40 45

Thr Lys Cys Ser Ser Thr Thr Cys Ala Ser Pro Pro Ala Gly Val Gly
50 55 60

Asn Pro Pro Leu Pro Pro Gly Pro Val Pro Trp Pro Val Val Gly Asn
65 70 75 80

Leu Pro Glu Met Leu Leu Asn Lys Pro Ala Phe Arg Trp Ile His Gln
85 90 95

Met Met Arg Glu Met Gly Thr Asp Ile Ala Cys Val Lys Leu Gly Gly
100 105 110

Val His Val Val Ser Ile Thr Cys Pro Glu Ile Ala Arg Glu Val Leu
115 120 125

Arg Lys Gln Asp Ala Asn Phe Ile Ser Arg Pro Leu Thr Phe Ala Ser
130 135 140

Glu Thr Phe Ser Gly Gly Tyr Arg Asn Ala Val Leu Ser Pro Tyr Gly
145 150 155 160

Asp Gln Trp Lys Lys Met Arg Arg Val Leu Thr Ser Glu Ile Ile Cys
165 170 175

Pro Ser Arg His Ala Trp Leu His Asp Lys Arg Thr Asp Glu Ala Asp
180 185 190

Asn Leu Thr Arg Tyr Val Tyr Asn Leu Ala Thr Lys Ala Ala Thr Gly
195 200 205

Asp Val Ala Val Asp Val Arg His Val Ala Arg His Tyr Cys Gly Asn
210 215 220

Val Ile Arg Arg Leu Met Phe Asn Arg Arg Tyr Phe Gly Glu Pro Gln

-continued

```
225                 230                 235                 240

Ala Asp Gly Gly Pro Gly Pro Met Glu Val Leu His Met Asp Ala Val
                245                 250                 255

Phe Thr Ser Leu Gly Leu Leu Tyr Ala Phe Cys Val Ser Asp Tyr Leu
            260                 265                 270

Pro Trp Leu Arg Gly Leu Asp Leu Asp Gly His Glu Lys Ile Val Lys
        275                 280                 285

Glu Ala Asn Val Ala Val Asn Arg Leu His Asp Thr Val Ile Asp Asp
    290                 295                 300

Arg Trp Arg Gln Trp Lys Ser Gly Glu Arg Gln Glu Met Glu Asp Phe
305                 310                 315                 320

Leu Asp Val Leu Ile Thr Leu Lys Asp Ala Gln Gly Asn Pro Leu Leu
                325                 330                 335

Thr Ile Glu Glu Val Lys Ala Gln Ser Gln Asp Ile Thr Phe Ala Ala
            340                 345                 350

Val Asp Asn Pro Ser Asn Ala Val Glu Trp Ala Leu Ala Glu Met Val
        355                 360                 365

Asn Asn Pro Glu Val Met Ala Lys Ala Met Glu Glu Leu Asp Arg Val
    370                 375                 380

Val Gly Arg Glu Arg Leu Val Gln Glu Ser Asp Ile Pro Lys Leu Asn
385                 390                 395                 400

Tyr Val Lys Ala Cys Ile Arg Glu Ala Phe Arg Leu His Pro Val Ala
                405                 410                 415

Pro Phe Asn Val Pro His Val Ala Leu Ala Asp Thr Thr Ile Ala Gly
            420                 425                 430

Tyr Arg Val Pro Lys Gly Ser His Val Ile Leu Ser Arg Thr Gly Leu
        435                 440                 445

Gly Arg Asn Pro Arg Val Trp Asp Glu Pro Leu Arg Phe Tyr Pro Asp
    450                 455                 460

Arg His Leu Ala Thr Ala Ala Ser Asp Val Ala Leu Thr Glu Asn Asp
465                 470                 475                 480

Leu Arg Phe Ile Ser Phe Ser Thr Gly Arg Arg Gly Cys Ile Ala Ala
                485                 490                 495

Ser Leu Gly Thr Ala Met Ser Val Met Leu Phe Gly Arg Leu Leu Gln
            500                 505                 510

Gly Phe Thr Trp Ser Lys Pro Ala Gly Val Glu Ala Val Asp Leu Ser
        515                 520                 525

Glu Ser Lys Ser Asp Thr Phe Met Ala Thr Pro Leu Val Leu His Ala
    530                 535                 540

Glu Pro Arg Leu Pro Ala His Leu Tyr Pro Ser Ile Ser Ile
545                 550                 555
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

-continued

```
Met Ala Thr Met Glu Val Glu Ala Ala Ala Ala Thr Val Leu Ala Ala
1               5                   10                  15

Pro
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Val Trp Asp Glu Pro Leu Arg
1               5
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Tyr Val Tyr Asn Leu Ala Thr Lys
1               5
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Ser Asp Thr Phe Met Ala Thr Pro Leu Val Ser Ser Ala Glu Pro Arg
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

-continued

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Ala Gln Ser Gln Asp Ile Thr Phe Ala Ala Val Asp Asn Pro Ser Asn
1               5                   10                  15

Ala Val Glu Xaa Ala Leu Ala Glu Met Val Asn Asn Pro Glu Val Met
            20                  25                  30

Ala Lys
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Ala Gln Gly Asn Pro Leu Leu Thr Ile Glu Glu Val Lys
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
Leu Val Gln Glu Ser Asp Ile Pro Lys
1               5
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Ile Ser Phe Ser Thr Gly
1               5
```

-continued

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Leu Pro Ala His Leu Tyr Pro Ser Ile Ser Leu His
1 5 10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Asp Leu Ala Asp Ile Pro Lys Gln Gln Arg Leu Met Ala Gly Asn
1 5 10 15

Ala Leu Val Val
20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Ala Arg Leu Ala Glu Ile Phe Ala Thr Ile Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

-continued

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Glu Asp Phe Thr Val Thr Thr Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Gln Tyr Ala Ala Leu Gly Ser Val Phe Thr Val Pro Ile Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Xaa Xaa Pro Phe Pro Ile
1 5

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Oligonucleotide specifying AA sequence MEVEAA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGGARGTNG ARGCNGC 17

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

-continued

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Oligonucleotide specifying AA sequence DFTMAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAYACNTTYA TGGCNAC 17

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: TYROL1b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CGGGATCCAT ATGCTGCTGT TATTAGCAGT TTTTCTGTCG TA 42

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: TYROL2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GACCGGCCGA AGCTTTAATT AGATGGAGAT GGA 33

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Tyrol3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AGTGGATCCA GCGGAATGCC GGCTT 25

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid

-continued

```
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

    (iii) HYPOTHETICAL: NO

    (vii) IMMEDIATE SOURCE:
          (B) CLONE: TYROL4

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CGTCATGCTC TTCGGAA                                                    17


(2) INFORMATION FOR SEQ ID NO: 23:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 51 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

    (iii) HYPOTHETICAL: NO

    (iii) ANTI-SENSE: NO

    (vii) IMMEDIATE SOURCE:
          (B) CLONE: TYROL1d

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CGGGATCCAT ATGGCTCTGT TATTAGCAGT TTTTCTGTCG TACCTGGCCC G              51


(2) INFORMATION FOR SEQ ID NO:24:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 13 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: other nucleic acid
          (A) DESCRIPTION: /desc = "oligonucleotide with an EcoRI site"

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCAGGAATTC CGG                                                        13
```

What is claimed is:

1. An isolated cytochrome P-450 monooxygenase that catalyzes the conversion of an amino acid to the corresponding N-hydroxyamino acid and the conversion of said N-hydroxyamino acid to the corresponding oxime.

2. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase is obtained from a plant that produces cyanogenic glycosides or glucosinolates.

3. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase is obtained from a plant selected from the group consisting of the genera Sorghum, Trifolium, Linum, Taxus, Triglochin, Mannihot, Amygdalus, Prunus, and cruciferous plants.

4. An isolated cytochrome P-450 monooxygenase according to claim **3**, wherein said monooxygenase is obtained from *Sorghum bicolor*.

5. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said amino acid is selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine, and cyclopentenylglycine isoleucine.

6. An isolated cytochrome P-450 monooxygenase according to claim **5**, wherein said amino acid is tyrosine.

7. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase has a molecular weight of 57 kD, as determined by SDS-PAGE.

8. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase comprises an N-terminal amino acid sequence as shown in SEQ ID NO: 3.

9. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase comprises an amino acid sequence as shown in SEQ ID NO: 2.

10. An isolated cytochrome P-450 monooxygenase according to claim **1**, wherein said monooxygenase comprises an internal amino acid sequence selected from the group consisting of SEQ ID NOs: 4–11.

11. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 4.

12. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 5.

13. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 6.

14. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 7.

15. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 8.

16. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 9.

17. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 10.

18. An isolated cytochrome P-450 monooxygenase according to claim **10**, wherein said monooxygenase comprises an internal amino acid sequence as shown in SEQ ID NO: 11.

19. A method of producing an isolated cytochrome P-450 monooxygenase according to claim **1**, comprising:

a) engineering a gene encoding said monooxygenase to be expressible in a host organism;

b) transforming said host organism with the engineered gene; and

c) isolating the monooxygenase from the host organism or a culture supernatant.

20. A method according to claim **19**, wherein the host organism is selected from the group consisting of bacteria, yeast, and insect cells.

21. An isolated cytochrome P-450 monooxygenase that catalyzes the conversion of an oxime to the corresponding nitrile and the conversion of said nitrile to the corresponding cyanohydrine.

22. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said monooxygenase is obtained from a plant that produces cyanogenic glycosides or glucosinolates.

23. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said monooxygenase is obtained from a plant selected from the group consisting of the genera Sorghum, Trifolium, Linum, Taxus, Triglochin, Mannihot, Amygdalus, Prunus, and cruciferous plants.

24. An isolated cytochrome P-450 monooxygenase according to claim **23**, wherein said monooxygenase is obtained from *Sorghum bicolor*.

25. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said oxime is obtained by the conversion of an amino acid to the corresponding N-hydroxyamino acid and the conversion of said N-hydroxyamino acid to said oxime by another cytochrome P-450 monooxygenase.

26. An isolated cytochrome P-450 monooxygenase according to claim **25**, wherein said amino acid is selected from the group consisting of tyrosine, phenylalanine, tryptophan, valine, leucine, isoleucine, and cyclopentenylglycine isoleucine.

27. An isolated cytochrome P-450 monooxygenase according to claim **26**, wherein said amino acid is tyrosine.

28. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein the ability of said monooxygenase to convert an oxime to the corresponding nitrile depends on the presence of NADPH and wherein this dependency can be overcome by the addition of reductants.

29. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said monooxygenase has a molecular weight of 51 kD, as determined by SDS-PAGE.

30. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said monooxygenase comprises an N-terminal amino acid sequence as shown in SEQ ID NO: 12.

31. An isolated cytochrome P-450 monooxygenase according to claim **21**, wherein said monooxygenase comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, and SEQ ID NO: 16.

32. An isolated cytochrome P-450 monooxygenase according to claim **31**, wherein said monooxygenase comprises an amino acid sequence as shown in SEQ ID NO: 13.

33. An isolated cytochrome P-450 monooxygenase according to claim **31**, wherein said monooxygenase comprises an amino acid sequence as shown in SEQ ID NO: 14.

34. An isolated cytochrome P-450 monooxygenase according to claim **31**, wherein said monooxygenase comprises an amino acid sequence as shown in SEQ ID NO: 15.

35. An isolated cytochrome P-450 monooxygenase according to claim **31**, wherein said monooxygenase comprises an amino acid sequence as shown in SEQ ID NO: 16.

36. A method of producing an isolated cytochrome P-450 monooxygenase according to claim **21**, comprising:

a) engineering a gene encoding said monooxygenase to be expressible in a host organism;

b) transforming said host organism with the engineered gene; and

c) isolating the monooxygenase from the host organism or a culture supernatant.

37. A method according to claim **36**, wherein the host organism is selected from the group consisting of bacteria, yeast, and insect cells.

* * * * *

US006130077A

United States Patent [19]

Yue et al.

[11] Patent Number: 6,130,077

[45] Date of Patent: Oct. 10, 2000

[54] **HUMAN CYTOCHROME P450**

[75] Inventors: **Henry Yue**, Sunnyvale; **Karl J. Guegler**, Menlo Park; **Neil C. Corley**, Mountain View; **Janice Au-Young**, Berkeley, all of Calif.

[73] Assignee: **Incyte Pharmaceuticals, Inc.**, Palo Alto, Calif.

[21] Appl. No.: **09/165,241**

[22] Filed: **Oct. 1, 1998**

[51] **Int. Cl.**[7] **C12N 9/02**

[52] **U.S. Cl.** **435/189**; 536/23.2; 435/320.1; 435/6; 435/252.3; 435/254.11

[58] **Field of Search** 435/189, 320.1, 435/6, 207.33; 536/23.2

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,886,157 3/1999 Guengerich et al. 435/189

OTHER PUBLICATIONS

GenBank Accession No. AA580329, Sep. 1997.
GenBank Accession No. AA393407, Apr. 1997.
GenBank Accession No. N78537, Mar. 1996.
GenBank Accession No. AA094256, Oct. 1996.
ExPASY Enzyme: EC 1.14.14.1 (Sep., 1998).
Prosite: PDOC00081 Cytochrome P450 cysteine heme–iron ligand signature (Nov., 1997).
Prints EP4501 E–Class P450 Group I signature (Mar., 1996).
Graham–Lorence, S. and J.A. Peterson, "P450s: Structural similarities and functional differences," *FASEB J.*, 10:206–214 (1996).
Nebert, D.W. and F.J. Gonzalez, "P450 Genes: Structure, Evolution, and Regulation," *Ann. Rev. Biochem.*, 56:945–993 (1987).
Dogra, S.C. et al., "Transcriptional Activation of Cytochrome P450 Genes by Different Classes of Chemical Inducers," *Clin. Exp. Pharmacol. Physiol.*, 25:1–9 (1998).
Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, Subfamily I (Dioxin–Inducible), Polypeptide 1; CYP1B1 (Apr., 1997).
Katsung, B.G., Ed., *Basic and Clinical Pharmacology*, Appleton & Lange, Norwalk, CT, pp. 48–59 (1995).
Morgan, E.T., "Regulation of Cytochromes P450 During Inflammation and Infection," *Drug Metab. Rev.*, 29(4):1129–1188 (1997).
OMIM *240300 Autoimmune Polyendocrinopathy–Candidiasis–Ectodermal Dystrophy (Jun., 1986).
Isselbacher, K.J. et al., *Harrison's Principles of Internal Medicine*, McGraw–Hill, Inc., New York, NY, pp. 1968–1970 (1994).
Takeyama, K. et al., "25–Hydroxyvitamin D_3 1α–Hydroxylase and Vitamin D Synthesis," *Science*, 227:1827–1830 (1997).
Kitanaka, S. et al., "Inactivating Mutations in the 25–Hydroxyvitamin D_3 1α–Hydroxylase Gene in Patients with Pseudovitamin D–Deficiency Rickets," *N. Engl. J. Med.*, 338(10):653–661 (1998).
OMIM *213700 Cerebrotendinous Xanthomatosis (Jun., 1986).
OMIM #122700 Coumarin Resistance (Jun., 1986).
Agarwal, V.R. et al., "Molecular Basis of Severe Gynecomastia Associated with Aromatase Expression in a Fibrolamellar Hepatocellular Carcinoma," *J. Clin. Endocrinol. Metab.*, 83(5):1797–1800 (1998).
Waxman, D.J., "P450–Catalyzed Steroid Hydroxylation: Assay and Product Indentification by Thin–Layer Chromatography," *Methods Enzymol.*, 206:462–476 (1991).

Primary Examiner—Ponnathapu Achutamurthy
Assistant Examiner—Kathleen Kerr
Attorney, Agent, or Firm—Incyte Pharmaceuticals, Inc.

[57] **ABSTRACT**

The invention provides a human cytochrome P450 (HUCYP) and polynucleotides which identify and encode HUCYP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of HUCYP.

11 Claims, 5 Drawing Sheets

```
                9          18          27          36          45          54
5' GCC GAT CCG AGA CGT GGC TCC CTG GGC GGC AGA ACC ATG TTG GAC TTC GCG ATC
                                                   M   L   D   F   A   I

               63          72          81          90          99         108
   TTC GCC GTT ACC TTC TTG CTG GCG TTG GTG GGA GCC GTG CTC TAC CTC TAT CCG
   F   A   V   T   F   L   L   A   L   V   G   A   V   L   Y   L   Y   P

              117         126         135         144         153         162
   GCT TCC AGA CAA GCT GCA GGA ATT CCA GGG ATT ACT CCA ACT GAA GAA AAA GAT
   A   S   R   Q   A   A   G   I   P   G   I   T   P   T   E   E   K   D

              171         180         189         198         207         216
   GGT AAT CTT CCA GAT ATT GTG AAT AGT GGA AGT TTG CAT GAG TTC CTG GTT AAT
   G   N   L   P   D   I   V   N   S   G   S   L   H   E   F   L   V   N

              225         234         243         252         261         270
   TTG CAT GAG AGA TAT GGG CCT GTG GTC TCC TTC TGG TTT GGC AGG CGC CTC GTG
   L   H   E   R   Y   G   P   V   V   S   F   W   F   G   R   R   L   V

              279         288         297         306         315         324
   GTT AGT TTG GGC ACT GTT GAT GTA CTG AAG CAG CAT ATC AAT CCC AAT AAG ACA
   V   S   L   G   T   V   D   V   L   K   Q   H   I   N   P   N   K   T

              333         342         351         360         369         378
   TTG GAC CCT TTT GAA ACC ATG CTG AAG TCA TTA TTA AGG TAT CAA TCT GGT GGT
   L   D   P   F   E   T   M   L   K   S   L   L   R   Y   Q   S   G   G
```

FIGURE 1A

```
        387         396         405         414         423         432
GGC AGT GTG AGT GAA AAC CAC ATG AGG AAA AAA TTG TAT GAA AAT GGT GTG ACT
G   S   V   S   E   N   H   M   R   K   K   L   Y   E   N   G   V   T

        441         450         459         468         477         486
GAT TCT CTG AAG AGT AAC TTT GCC CTC CTC CTA AAG CTT TCA GAA GAA TTA TTA
D   S   L   K   S   N   F   A   L   L   L   K   L   S   E   E   L   L

        495         504         513         522         531         540
GAT AAA TGG CTC TCC TAC CCA GAG ACC CAG CAC GTG CCC CTC AGC CAG CAT ATG
D   K   W   L   S   Y   P   E   T   Q   H   V   P   L   S   Q   H   M

        549         558         567         576         585         594
CTT GGT TTT GCT ATG AAG TCT GTT ACA CAG ATG GTA ATG GGT AGT ACA TTT GAA
L   G   F   A   M   K   S   V   T   Q   M   V   M   G   S   T   F   E

        603         612         621         630         639         648
GAT GAT CAG GAA GTC ATT CGC TTC CAG AAG AAT CAT GGC ACA GTT TGG TCT GAG
D   D   Q   E   V   I   R   F   Q   K   N   H   G   T   V   W   S   E

        657         666         675         684         693         702
ATT GGA AAA GGC TTT CTA GAT GGG TCA CTT GAT AAA AAC ATG ACT CGG AAA AAA
I   G   K   G   F   L   D   G   S   L   D   K   N   M   T   R   K   K

        711         720         729         738         747         756
CAA TAT GAA GAT GCC CTC ATG CAA CTG GAG TCT GTT TTA AGG AAC ATC ATA AAA
Q   Y   E   D   A   L   M   Q   L   E   S   V   L   R   N   I   I   K
```

FIGURE 1B

```
        765         774         783         792         801         810
GAA CGA AAA GGA AGG AAC TTC AGT CAA CAT ATT TTC ATT GAC TCC TTA GTA CAA
E   R   K   G   R   N   F   S   Q   H   I   F   I   D   S   L   V   Q

        819         828         837         846         855         864
GGG AAC CTT AAT GAC CAA CAG ATC CTA GAA GAC AGT ATG ATA TTT TCT CTG GCC
G   N   L   N   D   Q   Q   I   L   E   D   S   M   I   F   S   L   A

        873         882         891         900         909         918
AGT TGC ATA ATA ACT GCA AAA TTG TGT ACC TGG GCA ATC TGT TTT TTA ACC ACC
S   C   I   I   T   A   K   L   C   T   W   A   I   C   F   L   T   T

        927         936         945         954         963         972
TCT GAA GAA GTT CAA AAA AAA TTA TAT GAA GAG ATA AAC CAA GTT TTT GGA AAT
S   E   E   V   Q   K   K   L   Y   E   E   I   N   Q   V   F   G   N

        981         990         999        1008        1017        1026
GGT CCT GTT ACT CCA GAG AAA ATT GAG CAG CTC AGA TAT TGT CAG CAT GTG CTT
G   P   V   T   P   E   K   I   E   Q   L   R   Y   C   Q   H   V   L

       1035        1044        1053        1062        1071        1080
TGT GAA ACT GTT CGA ACT GCC AAA CTG ACT CCA GTT TCT GCC CAG CTT CAA GAT
C   E   T   V   R   T   A   K   L   T   P   V   S   A   Q   L   Q   D

       1089        1098        1107        1116        1125        1134
ATT GAA GGA AAA ATT GAC CGA TTT ATT ATT CCT AGA GAG ACC CTC GTC CTT TAT
I   E   G   K   I   D   R   F   I   I   P   R   E   T   L   V   L   Y
```

FIGURE 1C

```
      1143        1152        1161        1170        1179        1188
GCC CTT GGT GTG GTA CTT CAG GAT CCT AAT ACT TGG CCA TCT CCA CAC AAG TTT
A   L   G   V   V   L   Q   D   P   N   T   W   P   S   P   H   K   F

      1197        1206        1215        1224        1233        1242
GAT CCA GAT CGG TTT GAT GAT GAA TTA GTA ATG AAA ACT TTT TCC TCA CTT GGA
D   P   D   R   F   D   D   E   L   V   M   K   T   F   S   S   L   G

      1251        1260        1269        1278        1287        1296
TTC TCA GGC ACA CAG GAG TGT CCA GAG TTG AGG TTT GCA TAT ATG GTG ACC ACA
F   S   G   T   Q   E   C   P   E   L   R   F   A   Y   M   V   T   T

      1305        1314        1323        1332        1341        1350
GTA CTT CTT AGT GTA TTG GTG AAG AGA CTG CAC CTA CTT TCT GTG GAG GGA CAG
V   L   L   S   V   L   V   K   R   L   H   L   L   S   V   E   G   Q

      1359        1368        1377        1386        1395        1404
GTT ATT GAA ACA AAG TAT GAA CTG GTA ACA TCA TCA AGG GAA GAA GCT TGG ATC
V   I   E   T   K   Y   E   L   V   T   S   S   R   E   E   A   W   I

      1413        1422        1431        1440        1449        1458
ACT GTC TCA AAG AGA TAT TAA AAT TTT ATA CAT TTA AAA TCA TTG TTA AAT TGA
T   V   S   K   R   Y

      1467        1476        1485        1494        1503        1512
TTG AGG AAA ACA ACC ATT TAA AAA AAA TCT ATG TTG AAT CCT TTT ATA AAC CAG
```

FIGURE 1D

```
      1521        1530        1539        1548        1557        1566
TAT CAC TTT GTA ATA TAA ACA CCT ATT TGT ACT TAA TTT TGT AAA TTT GGA TTT

      1575        1584        1593        1602        1611        1620
TTA TAT ATC ATA TTT TCT TAA TTC ATT GTA CAC ATT TGA CTT ACT GCA CAG TAT

      1629        1638        1647
ATT GAT CAT TTT AAT GGG AAA AAA AAA A 3'
```

FIGURE 1E

HUMAN CYTOCHROME P450

FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of a human cytochrome P450 and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders.

BACKGROUND OF THE INVENTION

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochrome P450s, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochrome P450s are abundant in the liver, but also occur in other tissues. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; and Graham-Lorence, S. and Peterson, J. A. (1996) FASEB J. 10:206–214.)

Four hundred cytochrome P450s have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochrome P450s (PRINTS EP450I E-Class P450 Group I signature).

All cytochrome P450s use a heme cofactor and share structural attributes. Most cytochrome P450s are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochrome P450s. A ten amino acid signature sequence in this hemeiron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochrome P450s, a membrane-spanning region is usually found in the first 15–20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC0081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D. W. and Gonzalez, F. J. (1987) Ann. Rev. Biochem. 56:945–993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is

induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S. C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1–9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1). Cytochrome P450s are also involved in drug interactions, because the induction of a cytochrome P450 by one drug may affect the metabolism of another drug by the enzyme (Katzung, B. G. (1995) *Basic and Clinical Pharmacoloy*, Appleton & Lange, Norwalk Conn., pp. 48–59).

Cytochrome P450s are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E. T. (1997) Drug Metab. Rev. 29:1129–1188). Effects observed in vivo can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM *240300 Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochrome P450s have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K. J. et al. (1994) *Harrison's Principles of Internal Medicine*, McGraw-Hill, Inc. New York, N.Y., pp. 1968–1970; Takeyama, K. et al. (1997) Science 277:1827–1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653–661; OMIM *213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V. R. (1998) J. Clin. Endocrinol. Metab. 83:1797–1800).

The discovery of a new human cytochrome P450 and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders.

SUMMARY OF THE INVENTION

The invention is based on the discovery of a new human cytochrome P450 (HUCYP), the polynucleotides encoding HUCYP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders.

The invention features a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also provides an isolated and purified polynucleotide encoding the polypeptide compris-

ing the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2, and an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides an isolated and purified polynucleotide having a sequence complementary to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of HUCYP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of HUCYP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

FIGS. 1A, 1B, 1C, 1D, and 1E show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of HUCYP. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco Calif.).

Table 1 shows the tools, programs, and algorithms used to analyze HUCYP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the 14 art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"HUCYP" refers to the amino acid sequences of substantially purified HUCYP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to HUCYP, increases or prolongs the duration of the effect of HUCYP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HUCYP.

An "allelic variant" is an alternative form of the gene encoding HUCYP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to

allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HUCYP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as HUCYP or a polypeptide with at least one functional characteristic of HUCYP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HUCYP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HUCYP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HUCYP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HUCYP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of HUCYP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of HUCYP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to HUCYP, decreases the amount or the duration of the effect of the biological or immunological activity of HUCYP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HUCYP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HUCYP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HUCYP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HUCYP or fragments of HUCYP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.). "Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR kit (Perkin-Elmer, Norwalk Conn.) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison Wis.). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HUCYP, by northern analysis is indicative of the presence of nucleic acids encoding HUCYP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HUCYP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) Gene 73:237–244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626–645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" or "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of HUCYP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HUCYP.

The phrases "nucleic acid" or "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in

the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding HUCYP, or fragments thereof, or HUCYP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of HUCYP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to HUCYP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

The Invention

The invention is based on the discovery of a new human cytochrome P450 (HUCYP), the polynucleotides encoding HUCYP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders.

Nucleic acids encoding the HUCYP of the present invention were identified in Incyte Clone 991729H1from the colon cDNA library (COLNNOT11) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2812042H1 (OVARNOT10), 4502047H1 (BRAVTXT02), 2100169H1 (BRAITUT02), 991729R6 and 991729H1 (COLNNOT11), 038338R6 (HUVENOB01), 1996448H1 (BRSTTUT03), 687353H1 (UTRSNOT02), 1700412F6 (BLADTUT05), 4538247H1 (OVARNOT12), and 2381571F6 (ISLTNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in FIGS. 1A, 1B, 1C, 1D, and 1E. HUCYP is 462 amino acids in length and has three potential N-glycosylation sites at residues N94, N217, and N246; eleven potential casein kinase 2 phosphorylation sites as residues T36, S53, S116, S155, S183, T184, T293, T294, T442, S449, and S450; and seven potential protein kinase C phosphorylation sites at residues S134, T219, T281, T333, T336, S449, and S459. PFAM analysis indicates that HUCYP resembles P450 proteins from I32 through R46 1. PRINTS analysis indicates that HUCYP resembles E-class P450 Group I proteins from H62 through L81, from C285 through G311, from A367 through D391; resembles E-class P450 Group II proteins from Q265 through T293, from T294 through G311, and from T377 through E392; resembles E-class P450 Group IV proteins from E56 through V79, from L372 through D390, and from C409 through V427; resembles P450 superfamily proteins from T294 through N307; and resembles B-class P450 proteins from Q373 through R388. HUCYP shares conserved residues with cytochrome P450s in the potential heme-iron ligand region at F403, S404, G405, C409, and P410. C409 is the potential heme-binding residue. Like other eukaryotic cytochrome P450s, HUCYP contains a potentially membrane-spanning region in its first 15 to 20 residues. SPScan and HMM indicate that HUCYP has a potential signal peptide from M1 through about A18 or a potential transmembrane sequence from M1 through about Y23. A fragment of SEQ ID NO:2 from about nucleotide 109 to about nucleotide 153 is useful in hybridization or amplification technologies to identify SEQ ID NO:2 and to distinguish between SEQ ID NO:2 and a related sequence. Northern analysis shows the expression of this sequence in various libraries, 72% of which are associated with cancer and cell proliferation, 28% of which are associated with inflammation and immune response, 25% of which are reproductive tissues, 19% of which are nervous tissues, 19% of which are gastrointestinal tissues, and 16% of which are cardiovascular tissues. Of particular note is the expression of HUCYP in cancers of the liver, colon, uterus, testes, lung, breast, brain, and bladder, and in leukemia.

The invention also encompasses HUCYP variants. A preferred HUCYP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HUCYP amino acid sequence, and which contains at least one functional or structural characteristic of HUCYP.

The invention also encompasses polynucleotides which encode HUCYP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes HUCYP.

The invention also encompasses a variant of a polynucleotide sequence encoding HUCYP. In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HUCYP. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HUCYP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HUCYP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HUCYP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HUCYP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HUCYP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HUCYP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HUCYP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HUCYP and HUCYP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HUCYP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, or to a fragment of SEQ ID NO:2, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399–407; Kimmel, A. R. (1987) Methods Enzymol. 152:507–511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C., more preferably of at least about 37° C., and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C. in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C. in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C. in 250 mM NaCl, 25 mM trisodium

citrate, 1% SDS, 50% formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C., more preferably of at least about 42° C., and most preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C. in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland Ohio), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway N.J.), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg Md.). Preferably, sequence preparation is automated with machines such as the Robbins Hydra microdispenser (Robbins Scientific, Sunnyvale Calif.), Hamilton MICROLAB 2200 (Hamilton, Reno Nev.), Peltier Thermal Cycler 200 (PTC200; M J Research, Watertown Mass.) and the ABI CATALYST 800 (Perkin-Elmer). Sequencing is then carried out using the ABI 373 or 377 DNA sequencing systems (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale Calif.), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F. M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7; Meyers, R. A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., pp. 856–853.)

The nucleic acid sequences encoding HUCYP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318–322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111–119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J. D. et al. (1991) Nucleic Acids Res. 19:3055–306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto Calif.) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth Minn.) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68° C. to 72° C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HUCYP may be cloned in recombinant DNA molecules that direct expression of HUCYP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HUCYP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HUCYP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding HUCYP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M. H. et al. (1980) Nucl. Acids Res. Symp. Ser. (7):215–223, and Horn, T. et al. (1980) Nucl. Acids Symp. Ser. (7):225–232.) Alternatively, HUCYP itself or a fragment thereof may be

synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J. Y. et al. (1995) Science 269:202–204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of HUCYP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R. M. and F. Z. Regnier (1990) Methods Enzymol. 182:392–421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) *Proteins. Structures and Molecular Properties*, W H Freeman, New York N.Y.)

In order to express a biologically active HUCYP, the nucleotide sequences encoding HUCYP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HUCYP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HUCYP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HUCYP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125–162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HUCYP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Press, Plainview N.Y., ch. 4, 8, and 16–17; Ausubel, F. M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HUCYP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HUCYP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HUCYP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla Calif.) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HUCYP into the vector's multiple cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503–5509.) When large quantities of HUCYP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HUCYP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HUCYP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol. 153:516–54; and Scorer, C. A. et al. (1994) Bio/Technology 12:181–184.)

Plant systems may also be used for expression of HUCYP. Transcription of sequences encoding HUCYP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307–311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671–1680; Broglie, R. et al. (1984) Science 224:838–843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85–105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York N.Y., pp. 191–196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HUCYP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HUCYP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655–3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J. J. et al. (1997) Nat Genet. 15:345–355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HUCYP in cell

lines is preferred. For example, sequences encoding HUCYP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk^- or apr^- cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223–232; Lowy, I. et al. (1980) Cell 22:817–823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr confers resistance to methotrexate; neo confers resistance to the aminoglycosides, neomycin and G-418; and als or pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567–3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1–14.) Additional selectable genes have been described, e.g., trpB and hisD, which alter cellular requirements for metabolites. (See, e.g., Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047–8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C. A. (1995) Methods Mol. Biol. 55:121–131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HUCYP is inserted within a marker gene sequence, transformed cells containing sequences encoding HUCYP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HUCYP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HUCYP and that express HUCYP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA—DNA or DNA—RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HUCYP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HUCYP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St Paul Minn., Sect. IV; Coligan, J. E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York N.Y.; and Pound, J. D. (1998) *Immunochemical Protocols*, Humana Press, Totowa N.J.).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HUCYP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HUCYP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison Wis.), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HUCYP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HUCYP may be designed to contain signal sequences which direct secretion of HUCYP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HUCYP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HUCYP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HUCYP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin

(HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HUCYP encoding sequence and the heterologous protein sequence, so that HUCYP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HUCYP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of HUCYP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55–60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of HUCYP may be synthesized separately and then combined to produce the full length molecule.

Therapeutics

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HUCYP and cytochrome P450 proteins. In addition, the expression of HUCYP is closely associated with cancer and cell proliferation, inflammation and immune response, reproductive tissues, nervous tissues, gastrointestinal tissues, and cardiovascular tissues. Therefore, HUCYP appears to play a role in cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders. In the treatment of disorders associated with increased HUCYP expression or activity, it is desirable to decrease the expression or activity of HUCYP. In the treatment of disorders associated with decreased HUCYP expression or activity, it is desirable to increase the expression or activity of HUCYP.

Therefore, in one embodiment, HUCYP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HUCYP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a metabolic disorder such as Addison's disease, cystic fibrosis, diabetes, fatty hepatocirrhosis, galactosemia, goiter, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypothyroidism hyperlipidemia, hyperlipemia, lipid myopathies, obesity, lipodystrophies, and phenylketonuria, congenital adrenal hyperplasia, pseudovitamin D-deficiency rickets, cerebrotendinous xanthomatosis, and coumarin resistance.

In another embodiment, a vector capable of expressing HUCYP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HUCYP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HUCYP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HUCYP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HUCYP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HUCYP including, but not limited to, those listed above.

In a further embodiment, an antagonist of HUCYP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HUCYP. Such disorders may include, but are not limited to, cell proliferative, developmental, autoimmune/inflammatory, and metabolic disorders as listed above. In one aspect, an antibody which specifically binds HUCYP may be used directly as an antagonist or indirectly as a targeting or

delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HUCYP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HUCYP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HUCYP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HUCYP may be produced using methods which are generally known in the art. In particular, purified HUCYP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HUCYP. Antibodies to HUCYP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HUCYP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HUCYP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HUCYP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HUCYP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495–497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31–42; Cote, R. J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026–2030; and Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109–120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851–6855; Neuberger, M. S. et al. (1984) Nature 312:604–608; and Takeda, S. et al. (1985) Nature 314:452–454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HUCYP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D. R. (1991) Proc. Natl. Acad. Sci. 88:10134–10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833–3837; Winter, G. et al. (1991) Nature 349:293–299.)

Antibody fragments which contain specific binding sites for HUCYP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W. D. et al. (1989) Science 246:1275–1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HUCYP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HUCYP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HUCYP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HUCYP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HUCYP epitopes, represents the average affinity, or avidity, of the antibodies for HUCYP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HUCYP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} l/mole are preferred for use in immunoassays in which the HUCYP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 l/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HUCYP, preferably in active form, from the antibody (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington D.C.; Liddell, J. E. and Cryer, A. (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York N.Y.).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and

suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1–2 mg specific antibody/ml, preferably 5–10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of HUCYP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding HUCYP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HUCYP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HUCYP. Thus, complementary molecules or fragments may be used to modulate HUCYP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HUCYP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HUCYP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding HUCYP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HUCYP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HUCYP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions −10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J. E. et al. (1994) in Huber, B. E. and B. I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco N.Y., pp. 163–177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HUCYP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HUCYP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C. K. et al. (1997) Nature Biotechnology 15:462–466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HUCYP, antibodies to HUCYP, and mimetics, agonists, antagonists, or inhibitors of HUCYP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including,

but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing, Easton Pa.).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HUCYP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HUCYP or fragments thereof, antibodies of HUCYP, and agonists, antagonists or inhibitors of HUCYP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Diagnostics

In another embodiment, antibodies which specifically bind HUCYP may be used for the diagnosis of disorders characterized by expression of HUCYP, or in assays to monitor patients being treated with HUCYP or agonists, antagonists, or inhibitors of HUCYP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HUCYP include methods which utilize the antibody and a label to detect HUCYP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HUCYP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HUCYP expression. Normal or standard values for HUCYP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HUCYP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HUCYP expressed in subject samples, control and disease from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HUCYP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HUCYP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HUCYP, and to monitor regulation of HUCYP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HUCYP or closely related molecules may be used to identify nucleic acid sequences which encode HUCYP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding HUCYP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the HUCYP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the HUCYP gene.

Means for producing specific hybridization probes for DNAs encoding HUCYP include the cloning of polynucleotide sequences encoding HUCYP or HUCYP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or 35S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HUCYP may be used for the diagnosis of disorders associated with expression of HUCYP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a metabolic disorder such as Addison's disease, cystic fibrosis, diabetes, fatty hepatocirrhosis, galactosemia, goiter, hyperadrenalism,

hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypothyroidism hyperlipidemia, hyperlipemia, lipid myopathies, obesity, lipodystrophies, and phenylketonuria, congenital adrenal hyperplasia, pseudovitamin D-deficiency rickets, cerebrotendinous xanthomatosis, and coumarin resistance. The polynucleotide sequences encoding HUCYP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HUCYP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HUCYP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HUCYP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HUCYP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HUCYP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HUCYP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or over-expressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HUCYP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HUCYP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HUCYP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HUCYP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P. C. et al. (1993) J. Immunol. Methods 159:235–244; Duplaa, C. et al. (1993) Anal. Biochem. 229–236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614–10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150–2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HUCYP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J. J. et al. (1997) Nat Genet. 15:345–355; Price, C. M. (1993) Blood Rev. 7:127–134; and Trask, B. J. (1991) Trends Genet. 7:149–154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965–968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HUCYP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New

sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22–23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R. A. et al. (1988) Nature 336:577–580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HUCYP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HUCYP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HUCYP, or fragments thereof, and washed. Bound HUCYP is then detected by methods well known in the art. Purified HUCYP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HUCYP specifically compete with a test compound for binding HUCYP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HUCYP.

In additional embodiments, the nucleotide sequences which encode HUCYP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

I. cDNA Library Construction

The COLNNOT11 cDNA library was constructed from microscopically normal colon tissue obtained from a 60-year-old Caucasian male who had undergone a left hemicolectomy to remove grade 3 (of 4) adenocarcinoma in a different part of his bowel. The patient history reported previous diagnoses of depressive disorder and thrombophlebitis, accompanied by inflammatory polyarthropathies and inflammatory disease of the prostate.

The frozen tissue was homogenized and lysed using a Polytron PT-3000 homogenizer (Brinkmann Instruments, Westbury N.J.) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using a SW28 rotor in a L8-70M ultracentrifuge (Beckman Instruments, Fullerton Calif.) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.0, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water and treated with DNase at 37° C. The RNA extraction and precipitation were repeated as before. The mRNA was isolated using an OLIGOTEX kit (QIAGEN, Valencia Calif.) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERSCRIPT plasmid system (Life Technologies). cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech), and those cDNAs exceeding 400 bp were ligated into pSPORT1 plasmid (Life Technologies). The plasmid pSport I was subsequently transformed into DH5α competent cells (Life Technologies).

II. Isolation of cDNA Clones

Plasmid DNA was released from the cells and purified using the R.E.A.L. PREP 96 plasmid kit from QIAGEN. This kit enables the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Life Technologies) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using the ABI CATALYST 800 (Perkin-Elmer) or the HYDRA microdispenser (Robbins Scientific) or MICROLAB 2200 (Hamilton) systems in combination with PTC-200 thermal cyclers (MJ Research). The cDNAs were sequenced using the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) and standard ABI protocols, base calling software, and kits. In one alternative, cDNAs were sequenced using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics). In another alternative, the cDNAs were amplified and sequenced using the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). In yet another alternative, cDNAs were sequenced using solutions and dyes from Amersham Pharmacia Biotech. Reading frames for the ESTs were determined using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the tools, programs, and algorithms used and provides descriptions, references, and threshold parameters when applicable. The first column of Table 1 shows the tools, programs, and algorithms used, the second column provides brief descriptions, the third column presents the appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the

greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases, such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases, SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were used to identify polynucleotide sequence fragments from SEQ ID NO:2.

Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in the Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as the GenBank or LIFESEQ databases (Incyte Pharmaceuticals, Palo Alto Calif.). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses reported a percentage distribution of libraries in which the transcript encoding HUCYP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease categories included cancer, inflammation/trauma, fetal, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease expression are reported in the description of the invention.

V. Extension of HUCYP Encoding Polynucleotides

The full length nucleic acid sequence of SEQ ID NO:2 was produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68° C. to about 72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94° C., 3 min; Step 2: 94° C., 15 sec; Step 3: 60° C., 1 min; Step 4: 68° C., 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68° C., 5 min; Step 7: storage at 4° C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94° C., 3 min; Step 2: 94° C., 15 sec; Step 3: 57° C., 1 min; Step 4: 68° C., 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68° C., 5 min; Step 7: storage at 4° C.

The concentration of DNA in each well was determined by dispensing 100 μl PICO GREEN quantitation reagent (0.25% (v/v) PICO GREEN; Molecular Probes, Eugene Org.) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton Mass.), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison Wis.), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly Mass.) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA

polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37° C. in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94° C., 3 min; Step 2: 94° C., 15 sec; Step 3: 60° C., I min; Step 4: 72° C., 2 min; Step 5: step 2, 3, and 4 repeated 29 times; Step 6: 72° C., 5 min; Step 7: storage at 4° C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequence of SEQ ID NO:2 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [^{32}P]-adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston Mass.). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xbal, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham N.H.). Hybridization is carried out for 16 hours at 40° C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1×saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized and compared using autoradiography.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467–470; Shalon, D. et al. (1996) Genome Res. 6:639–645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the HUCYP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HUCYP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HUCYP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HUCYP-encoding transcript.

IX. Expression of HUCYP

Expression and purification of HUCYP are achieved using bacterial or virus-based expression systems. For expression of HUCYP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the trp-lac (tac) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the lac operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HUCYP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HUCYP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HUCYP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224–25 3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937–1945.)

In most expression systems, HUCYP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HUCYP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified HUCYP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of HUCYP Activity

HUCYP activity is demonstrated as the ability to hydroxylate the steroid testosterone (Waxman, D. J. (1991) Methods Enzymol. 206:462–476). 4-^{14}C-testosterone is diluted with unlabeled testosterone dissolved in toluene to give a working solution of 5–7 mCi/mmol. Ten nanomoles of the working solution of 4-^{14}C-testosterone (final assay concentration 50 μM) are aliquoted to a 13×100 mm assay tube, and the solvent is evaporated under N_2. HUCYP, 0.1 M HEPES buffer, pH 7.4, and 0.1 mM EDTA are added to the tube, on ice, to give a final volume of 175 μl. A control tube is prepared that lacks HUCYP. Samples are vortexed briefly, and the tubes are transferred to a shaking 37° C. water bath. Reactions are initiated 4 minutes later by the addition of NADPH dissolved in 25 μl 0.1 M HEPES buffer, pH 7.4 to give a final NADPH concentration of 0.3 mM. Reactions are terminated after 10–20 minutes by the addition of 1 ml ethyl acetate and vortexed for 30 seconds. The hydroxylated products are separated from testosterone by thin layer chromatography and detected by autoradiography. Product migrations are compared with those of testosterone and of hydroxylated steroid standards to identify the products. Reaction products are cut out of the thin layer chromatography plate and counted in a liquid scintillation counter. The amount of hydroxylated steroid products found is proportional to HUCYP activity.

XI. Functional Assays

HUCYP function is assessed by expressing the sequences encoding HUCYP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad Calif.), both of which contain the cytomegalovirus promoter. 5–10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1–2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate cellular properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) *Flow Cytometry, Oxford*, New York N.Y.

The influence of HUCYP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HUCYP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success N.Y.). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HUCYP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of HUCYP Specific Antibodies

HUCYP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M. G. (1990) Methods Enzymol. 182:488–495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HUCYP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide Synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis Mo.) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring HUCYP Using Specific Antibodies

Naturally occurring or recombinant HUCYP is substantially purified by immunoaffinity chromatography using antibodies specific for HUCYP. An immunoaffinity column is constructed by covalently coupling anti-HUCYP antibody to an activated chromatographic resin, such as CNBr-

activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HUCYP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HUCYP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HUCYP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HUCYP is collected.

XIV. Identification of Molecules Which Interact with HUCYP

HUCYP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529–539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HUCYP, washed, and any wells with labeled HUCYP complex are assayed. Data obtained using different concentrations of HUCYP are used to calculate values for the number, affinity, and association of HUCYP with the candidate molecules. Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 2

<210> SEQ ID NO 1
<211> LENGTH: 462
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE: -
<223> OTHER INFORMATION: 991729

<400> SEQUENCE: 1

```
Met Leu Asp Phe Ala Ile Phe Ala Val Thr Phe Leu Leu Ala Leu
  1               5                  10                  15

Val Gly Ala Val Leu Tyr Leu Tyr Pro Ala Ser Arg Gln Ala Ala
             20                  25                  30

Gly Ile Pro Gly Ile Thr Pro Thr Glu Glu Lys Asp Gly Asn Leu
             35                  40                  45

Pro Asp Ile Val Asn Ser Gly Ser Leu His Glu Phe Leu Val Asn
             50                  55                  60

Leu His Glu Arg Tyr Gly Pro Val Val Ser Phe Trp Phe Gly Arg
             65                  70                  75

Arg Leu Val Val Ser Leu Gly Thr Val Asp Val Leu Lys Gln His
             80                  85                  90

Ile Asn Pro Asn Lys Thr Leu Asp Pro Phe Glu Thr Met Leu Lys
             95                 100                 105

Ser Leu Leu Arg Tyr Gln Ser Gly Gly Gly Ser Val Ser Glu Asn
            110                 115                 120

His Met Arg Lys Lys Leu Tyr Glu Asn Gly Val Thr Asp Ser Leu
            125                 130                 135

Lys Ser Asn Phe Ala Leu Leu Leu Lys Leu Ser Glu Glu Leu Leu
            140                 145                 150

Asp Lys Trp Leu Ser Tyr Pro Glu Thr Gln His Val Pro Leu Ser
            155                 160                 165

Gln His Met Leu Gly Phe Ala Met Lys Ser Val Thr Gln Met Val
            170                 175                 180

Met Gly Ser Thr Phe Glu Asp Asp Gln Glu Val Ile Arg Phe Gln
            185                 190                 195

Lys Asn His Gly Thr Val Trp Ser Glu Ile Gly Lys Gly Phe Leu
            200                 205                 210
```

-continued

```
Asp Gly Ser Leu Asp Lys Asn Met Thr Arg Lys Lys Gln Tyr Glu
                215                 220                 225

Asp Ala Leu Met Gln Leu Glu Ser Val Leu Arg Asn Ile Ile Lys
                230                 235                 240

Glu Arg Lys Gly Arg Asn Phe Ser Gln His Ile Phe Ile Asp Ser
                245                 250                 255

Leu Val Gln Gly Asn Leu Asn Asp Gln Gln Ile Leu Glu Asp Ser
                260                 265                 270

Met Ile Phe Ser Leu Ala Ser Cys Ile Ile Thr Ala Lys Leu Cys
                275                 280                 285

Thr Trp Ala Ile Cys Phe Leu Thr Thr Ser Glu Glu Val Gln Lys
                290                 295                 300

Lys Leu Tyr Glu Glu Ile Asn Gln Val Phe Gly Asn Gly Pro Val
                305                 310                 315

Thr Pro Glu Lys Ile Glu Gln Leu Arg Tyr Cys Gln His Val Leu
                320                 325                 330

Cys Glu Thr Val Arg Thr Ala Lys Leu Thr Pro Val Ser Ala Gln
                335                 340                 345

Leu Gln Asp Ile Glu Gly Lys Ile Asp Arg Phe Ile Ile Pro Arg
                350                 355                 360

Glu Thr Leu Val Leu Tyr Ala Leu Gly Val Val Leu Gln Asp Pro
                365                 370                 375

Asn Thr Trp Pro Ser Pro His Lys Phe Asp Pro Asp Arg Phe Asp
                380                 385                 390

Asp Glu Leu Val Met Lys Thr Phe Ser Ser Leu Gly Phe Ser Gly
                395                 400                 405

Thr Gln Glu Cys Pro Glu Leu Arg Phe Ala Tyr Met Val Thr Thr
                410                 415                 420

Val Leu Leu Ser Val Leu Val Lys Arg Leu His Leu Leu Ser Val
                425                 430                 435

Glu Gly Gln Val Ile Glu Thr Lys Tyr Glu Leu Val Thr Ser Ser
                440                 445                 450

Arg Glu Glu Ala Trp Ile Thr Val Ser Lys Arg Tyr
                455                 460
```

<210> SEQ ID NO 2
<211> LENGTH: 1648
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: -
<223> OTHER INFORMATION: 991729

<400> SEQUENCE: 2

```
gccgatccga gacgtggctc cctgggcggc agaaccatgt tggacttcgc gatcttcgcc     60

gttaccttct tgctggcgtt ggtgggagcc gtgctctacc tctatccggc ttccagacaa    120

gctgcaggaa ttccagggat tactccaact gaagaaaaag atggtaatct tccagatatt    180

gtgaatagtg gaagtttgca tgagttcctg gttaatttgc atgagagata tgggcctgtg    240

gtctccttct ggtttggcag gcgcctcgtg gttagtttgg gcactgttga tgtactgaag    300

cagcatatca atcccaataa gacattggac ccttttgaaa ccatgctgaa gtcattatta    360

aggtatcaat ctggtggtgg cagtgtgagt gaaaaccaca tgaggaaaaa attgtatgaa    420

aatggtgtga ctgattctct gaagagtaac tttgccctcc tcctaaagct ttcagaagaa    480

ttattagata aatggctctc ctacccagag acccagcacg tgcccctcag ccagcatatg    540
```

-continued

```
cttggttttg ctatgaagtc tgttacacag atggtaatgg gtagtacatt tgaagatgat    600

caggaagtca ttcgcttcca gaagaatcat ggcacagttt ggtctgagat tggaaaaggc    660

tttctagatg ggtcacttga taaaaacatg actcggaaaa aacaatatga agatgccctc    720

atgcaactgg agtctgtttt aaggaacatc ataaaagaac gaaaaggaag gaacttcagt    780

caacatattt tcattgactc cttagtacaa gggaacctta atgaccaaca gatcctagaa    840

gacagtatga tattttctct ggccagttgc ataataactg caaaattgtg tacctgggca    900

atctgttttt taaccacctc tgaagaagtt caaaaaaaat tatatgaaga gataaaccaa    960

gtttttggaa atggtcctgt tactccagag aaaattgagc agctcagata ttgtcagcat   1020

gtgctttgtg aaactgttcg aactgccaaa ctgactccag tttctgccca gcttcaagat   1080

attgaaggaa aaattgaccg atttattatt cctagagaga ccctcgtcct ttatgccctt   1140

ggtgtggtac ttcaggatcc taatacttgg ccatctccac acaagtttga tccagatcgg   1200

tttgatgatg aattagtaat gaaaactttt tcctcacttg gattctcagg cacacaggag   1260

tgtccagagt tgaggtttgc atatatggtg accacagtac ttcttagtgt attggtgaag   1320

agactgcacc tactttctgt ggagggacag gttattgaaa caaagtatga actggtaaca   1380

tcatcaaggg aagaagcttg gatcactgtc tcaaagagat attaaaattt tatacattta   1440

aaatcattgt taaattgatt gaggaaaaca accatttaaa aaaaatctat gttgaatcct   1500

tttataaacc agtatcactt tgtaatataa acacctattt gtacttaatt ttgtaaattt   1560

ggatttttat atatcatatt ttcttaattc attgtacaca tttgacttac tgcacagtat   1620

attgatcatt ttaatgggaa aaaaaaaa                                      1648
```

What is claimed is:

1. An isolated and purified polynucleotide encoding a polypeptide

(a) comprising SEQ ID NO:1; or

(b) consisting of a fragment of SEQ ID NO:1 from about amino acid residue Ala at position 25 to about amino acid residue Glu at position 39 of SEQ ID NO:1.

2. An isolated and purified polynucleotide having at least 85% polynucleotide sequence identity to the polynucleotide of claim 1(*a*) over the full length of the sequence.

3. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 1.

4. An isolated and purified polynucleotide

(a) comprising the polynucleotide sequence of SEQ ID NO:2; or

(b) consisting of a fragment of SEQ ID NO:2 from about nucleotide 109 to about nucleotide 153 of SEQ ID NO:2.

5. An isolated and purified polynucleotide having at least 85% polynucleotide sequence identity to the polynucleotide of claim 4(*a*) over the full length of the sequence.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 4.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell comprising the expression vector of claim 7.

9. A method for producing a polypeptide, the method comprising the steps of:

(a) culturing the host cell of claim 8 under conditions suitable for the expression of the polypeptide; and

(b) recovering the polypeptide from the host cell culture.

10. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 3 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

11. The method of claim 10 further comprising amplifying the polynucleotide prior to hybridization.

* * * * *

US005801024A

United States Patent [19]

van den Brink et al.

[11] Patent Number: 5,801,024

[45] Date of Patent: Sep. 1, 1998

[54] **OXIDOREDUCTASE FROM FILAMENTOUS FUNGI, DNA CODING THEREFOR AND CELLS TRANSFORMED WITH SAID DNA**

[75] Inventors: **Johannes Maarten van den Brink**, Utrecht; **Robertus Franciscus van Gorcom**, Delft, both of Netherlands

[73] Assignee: **Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek TNO**, Netherlands

[21] Appl. No.: **553,279**

[22] PCT Filed: **Jun. 10, 1994**

[86] PCT No.: **PCT/NL94/00135**

§ 371 Date: **Nov. 28, 1995**

§ 102(e) Date: **Nov. 28, 1995**

[87] PCT Pub. No.: **WO94/29453**

PCT Pub. Date: **Dec. 22, 1994**

[30] **Foreign Application Priority Data**

Jun. 11, 1993 [NL] Netherlands 9301025

[51] **Int. Cl.**[6] **C12P 7/00**; C12P 7/40; C12N 9/02; C07H 21/04

[52] **U.S. Cl.** **435/132**; 435/136; 435/189; 435/190; 435/191; 435/254.11; 435/254.3; 435/252.3; 435/320.1; 435/325; 536/23.2; 935/22

[58] **Field of Search** 435/189, 190, 435/191, 132, 136, 320.1, 254.3, 254.11, 325; 536/23.2, 23.7, 23.74; 935/22

[56] **References Cited**

PUBLICATIONS

van Gorcom et al. (1990) Isolation and molecular characterisation of the benzoate–para–hydroxylase gene (bphA) of Aspergillus niger: A member of a new gene family of the cytochrome P450 superfamily, Mol. Gen. Genet. 223: 192–197, Dec. 1990.

Ngo et al. (1994) Computational Complexity, Protein Structure Preticition, and the Levinthal Paradox, In The Protein Folding Problem and Tertiary Structure Prediction, Eds. Merz et al., Birkhauser, Boston, MA, pp. 491–495, Jan. 1994.

van den Brink et al. (1995) Cloning and Characterization of the NADPH Cytochrome P450 Oxidoreductase Gene from the Filamentous Fungus Aspergillus niger, DNA and Cell Biology 8(14):719–729, Aug. 1995.

Primary Examiner—Robert A. Wax
Assistant Examiner—Einar Stole
Attorney, Agent, or Firm—Londa and Traub LLP

[57] **ABSTRACT**

Disclosed is a new gene which codes for a NADPH cytochrome P450 oxidoreductase from filamentous fungi. Also disclosed are a recombinant DNA molecule which comprises at least a part of that gene, an RNA molecule derived therefrom and a new polypeptide (or protein), as well as at least one host cell transformed with such recombinant DNA molecule. Particularly suitable host cells are those originating from a filamentous fungus. With the polypeptides (proteins) and the host cells according to the invention, new processes for enzymatic conversions can be performed. In particular, enzymatic conversions by means of monooxygenases, more particularly by means of enzymes from the cytochrome P450 superfamily are improved with the invention.

13 Claims, 5 Drawing Sheets

```
TGATAACTCC TCAGCAAATC GGAGTAAACA GAAGGACAAG TCATTGGAGT ACTAAGTAGC      60
TCCGTGTCAG AGACCCGGAC AGGATCAGCT TCTCCGAACC CGAGACTCCG GGCGAAAAGG     120
CCACCATCGC TCAGGCTACC ACCTGTGTTC CTTCCGTCGA TCGTCCTCCC TCGTTTCCGG     180
CTCACGGCCC CCCAAATTAT TGCGGTCTGC TTAGCAGTGG GTTCGGCCTC TCTGTTCTTC     240
CTGGATCACA CCACGGCTTA CTTTCTTATC CTTTTCCTTT TCCTTTCTTC CTTTCTTCCT     300
GTTCTCCTTT CTTCCTTTCC ACCCCCTTCT TTCTTTTAAC CCCATAGCGT CATTCTTTCT     360
TCCGTTTTAT CTTTGGTTTT GGGACGCCGC CACCTTATCT CGGTTCCTGC CTCGGTCTCC     420
GGTGATCGCA CCTGGATAGG CTAAGCGTAG GGAGGTGTGA CATTCTTCTT TCACCTCCTC     480
TCCTTTTCCC GCCTCACTCC GTTCAATCCC CCGCTCCACC CTTTCAGACT CGCCATCGTA     540
TCAAGTCGGG GCCTTTGCTT GCGCCGCTGA ACAGCCTCAC CATGGCGCAA CTCGATACCC     600
TCGATCTGGT GGTCCTGGCG GTGCTTTTGG TGGGTAGCGT GGCCTACTTC ACCAAGGGCA     660
CCTACTGGGC AGTTGCAAAG ACCCGTATGC CTCTACCGGC CCCGCGGATG AACGGCGCCG     720
CTAAGGCTGG CAAGACTCGG AACATCATTG AGAAGATGGA AGAAACGGGC AAGAATTGTG     780
TTATTTTCTA CGGATCGCAA ACTGGAACCG CTGAGGACTA CGCCTCCAGA TTGGCCAAGG     840
AAGGATCTCA GCGCTTCGGC CTCAAGACCA TGGTGGCTGA CCTCGAGGAA TACGACTATG     900
AGAACCTGGA CCAATTCCCG GAGGACAAGG TTGCGTTTTT CGTGCTCGCC ACCTACGGAG     960
AGGGTGAGCC TACGGATAAT GCTGTTGAGT TCTACCAGTT CTTCACCGGT GACGACGTTG    1020
CTTTTGAGAG CGCGTCCGCG GACGAGAAGC CTCTGTCCAA GCTGAAGTAT GTTGCTTTCG    1080
GTCTGGGTAA CAACACTTAT GAGCACTACA ACGCCATGGT TCGTCAAGTC GATGCTGCTT    1140
TCCAGAAGCT CGGGCCGCAG CGTATTGGTT CTGCTGGCGA GGGTGATGAC GGTGCCGGTA    1200
CAATGGAAGA AGACTTCTTG GCCTGGAAGG AGCCCATGTG GGCAGCACTG TCGGAGTCGA    1260
TGGATCTCGA AGAGCGTGAA GCGGTCTACG AACCTGTTTT CTGCGTCACC GAAAACGAGT    1320
CCCTGAGCCC TGAGGACGAG ACGGTCTATC TTGGAGAGCC CACCCAGAGC CACCTTCAGG    1380
GTACTCCCAA AGGCCCGTAC TCTGCGCACA ACCCCTTTAT CGCCCCTATT GCCGAATCTC    1440
GTGAGCTTTT CACCGTCAAG GATCGCAACT GTCTGCACAT GGAAATTAGC ATCGCTGGAA    1500
GTAACTTGTC CTACCAGACT GGTGACCACA TCGCTGTTTG GCCCACAAAC GCTGGTGCCG    1560
AAGTGGATCG GTTCCTTCAG GTCTTCGGTC TCGAGGGCAA GCGTGATTCG GTCATCAACA    1620
TCAAGGGTAT CGATGTTACG GCCAAGGTCC CAATCCCGAC CCCGACCACG TACGATGCCG    1680
CTGTTCGGTA CTATATGGAA GTCTGCGCCC CTGTGTCCCG TCAGTTTGTA GCCACTCTGG    1740
CCGCGTTCGC TCCGATGAGG AAAGCAAGGC AGAGATTGTG CGTCTGGGTA GCACAAGGAC    1800
TATTTCCACG AGAAGGTCAC CAACCAATGC TTCAACATGC CCAGGCTCTT CAGAGCATCA    1860
CGTCCAAGCC TTTCTCTGCT GTTCCGTTCT CTCTGCTTAT TGAAGGCATT ACGAAGCTGC    1920
AGCCTCGCTA CTACTCGATC TCTTCGTCCT CCCTTGTCCA GAAGGACAAG ATCAGCATCA    1980
CGGCCGTTGT GGAATCTGTT CGTCTGCCCG GTGCCTCTCA CATGGTGAAG GGTGTGACTA    2040
```

FIG.1

```
Thr Ala Lys Val Pro Ile Pro Thr Pro Thr Thr Tyr Asp Ala Ala Val
        355                 360                 365

Arg Tyr Tyr Met Glu Val Cys Ala Pro Val Ser Arg Gln Phe Val Ala
    370                 375                 380

Thr Leu Ala Ala Phe Ala Pro Met Arg Lys Ala Arg Gln Arg Leu Cys
385                 390                 395                 400

Val Trp Val Ala Gln Gly Leu Phe Pro Arg Glu Gly His Gln Pro Met
                405                 410                 415

Leu Gln His Ala Gln Ala Leu Gln Ser Ile Thr Ser Lys Pro Phe Ser
            420                 425                 430

Ala Val Pro Phe Ser Leu Leu Ile Glu Gly Ile Thr Lys Leu Gln Pro
        435                 440                 445

Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Val Gln Lys Asp Lys Ile
    450                 455                 460

Ser Ile Thr Ala Val Val Glu Ser Val Arg Leu Pro Gly Ala Ser His
465                 470                 475                 480

Met Val Lys Gly Val Thr Thr Asn Tyr Leu Leu Ala Leu Lys Gln Lys
                485                 490                 495

Gln Asn Gly Arg Ser Leu Ser Arg Pro Ser Arg Leu Asp Leu Leu His
            500                 505                 510

His Gly Pro Arg Asn Lys Tyr Asp Gly Ile His Val Pro Val His Val
        515                 520                 525

Arg His Ser Asn Phe Lys Leu Pro Ser Asp Pro Ser Arg Pro Ile Ile
    530                 535                 540

Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Arg Gly Phe Ile Gln
545                 550                 555                 560

Glu Arg Ala Ala Leu Ala Ala Lys Gly Glu Lys Val Gly Pro Thr Val
                565                 570                 575

Leu Phe Phe Gly Cys Arg Lys Ser Asp Glu Asp Phe Leu Tyr Lys Asp
            580                 585                 590

Glu Trp Lys Thr Tyr Gln Asp Gln Leu Gly Asp Asn Leu Lys Ile Ile
        595                 600                 605

Thr Ala Phe Ser Arg Glu Gly Pro Gln Lys Val Tyr Val Gln His Arg
    610                 615                 620

Leu Arg Glu His Ser Glu Leu Val Ser Asp Leu Leu Lys Gln Lys Ala
625                 630                 635                 640

Thr Phe Tyr Val Cys Gly Asp Ala Ala Asn Met Ala Arg Glu Val Asn
                645                 650                 655

Leu Val Leu Gly Gln Ile Ile Ala Ala Gln Arg Gly Leu Pro Ala Glu
            660                 665                 670

Lys Gly Glu Glu Met Val Lys His Met Arg Arg Arg Gly Arg Tyr Gln
        675                 680                 685

Glu Asp Val Trp Ser
    690
```

FIG. 2

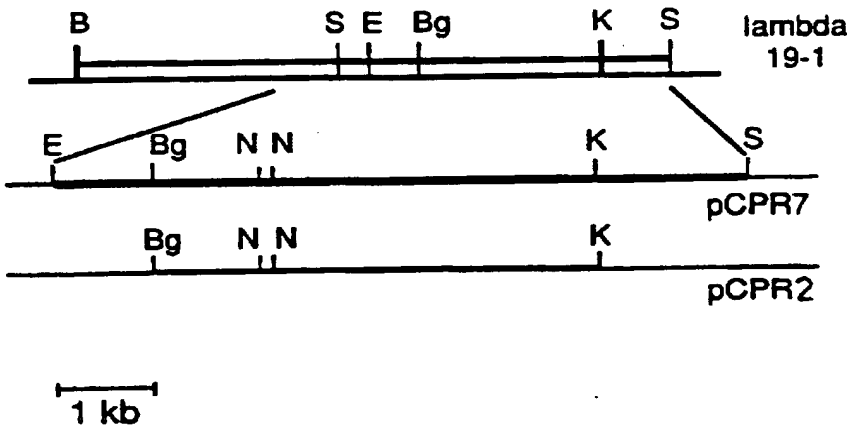
B
S E Bg
K S
lambda
19-1
E
Bg
N N
K
S
pCPR7
Bg
N N
K
pCPR2
1 kb

FIG. 3

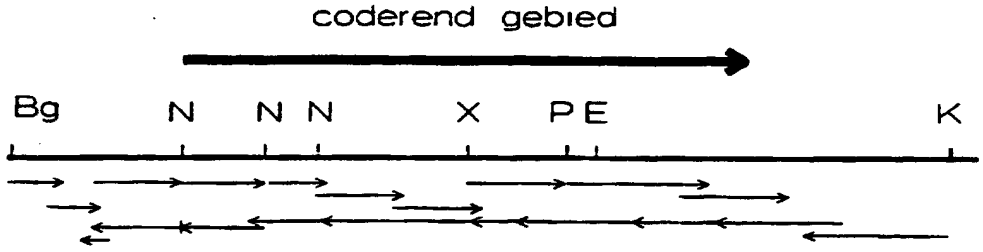
coderend gebied
Bg
N
N N
X
P E
K

FIG. 4

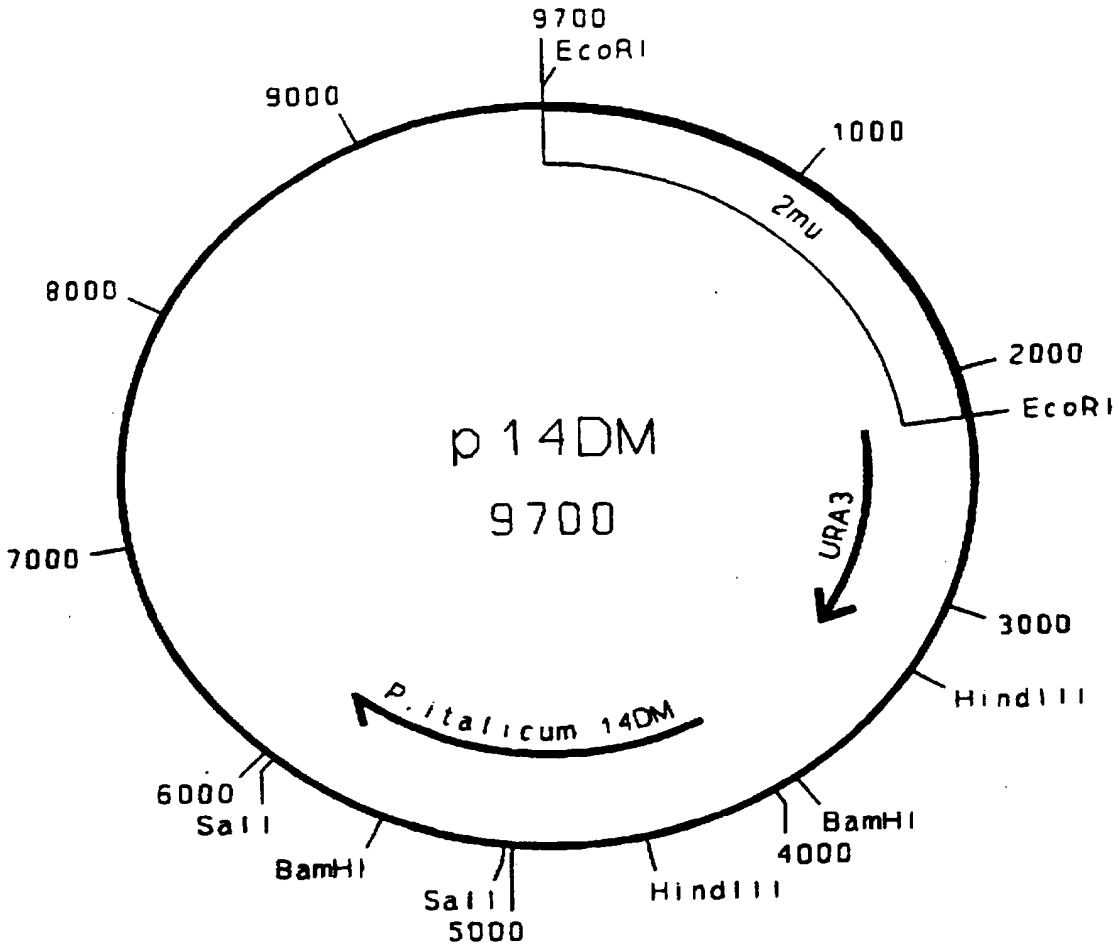
9700
EcoRI
1000
2000
EcoRI
3000
HindIII
BamHI
4000
HindIII
Sal I
5000
BamHI
Sal I
6000
7000
8000
9000
2mu
URA3
P. italicum 14DM
p14DM
9700

FIG. 5

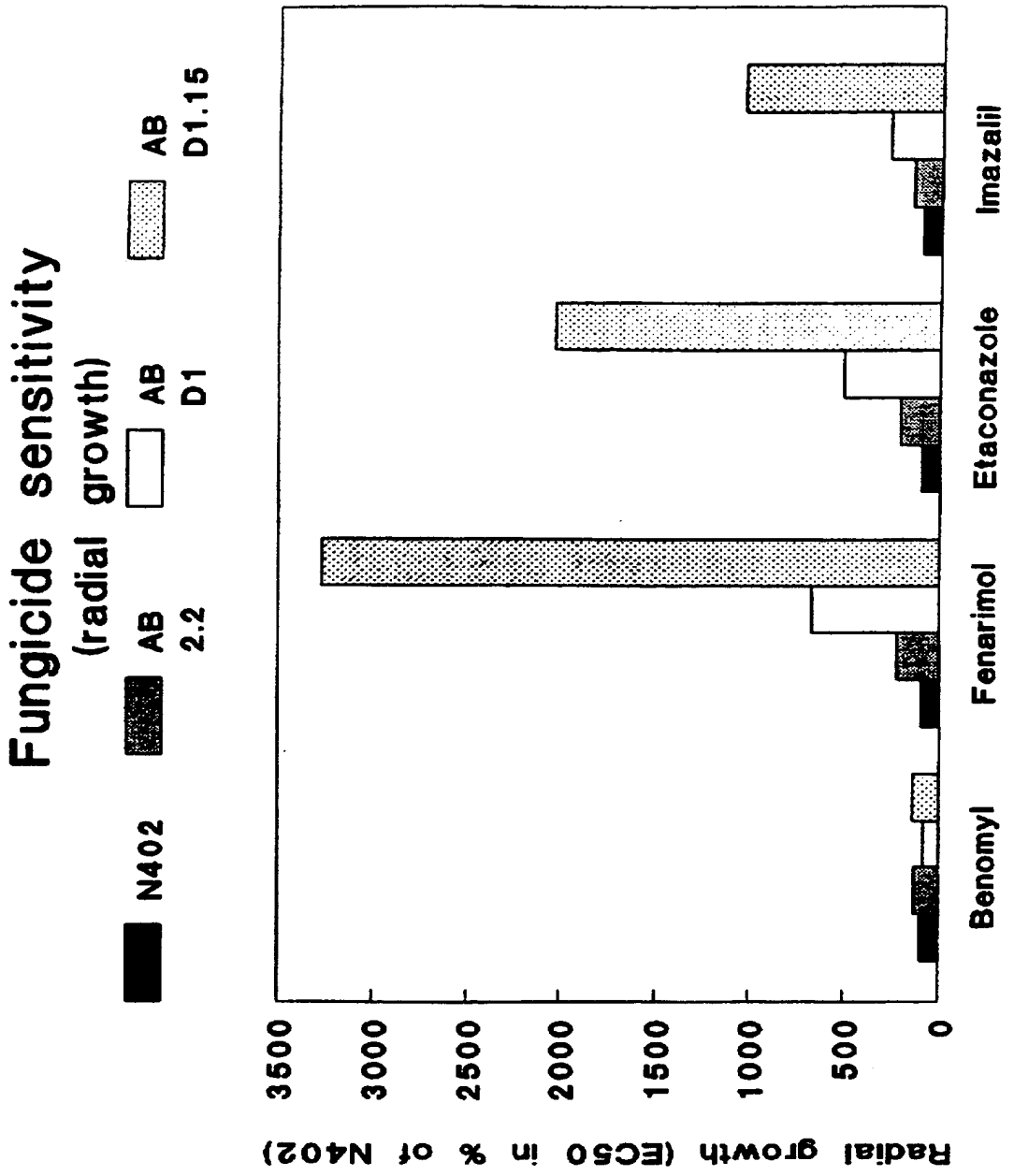
Fungicide sensitivity
(radial growth)
N402
AB 2.2
AB D1
AB D1.15
Radial growth (EC50 in % of N402)
3500
3000
2500
2000
1500
1000
500
0
Benomyl
Fenarimol
Etaconazole
Imazalil
Fungicide

FIG.6

OXIDOREDUCTASE FROM FILAMENTOUS FUNGI, DNA CODING THEREFOR AND CELLS TRANSFORMED WITH SAID DNA

BACKGROUND OF THE INVENTION

The invention relates to a new gene, a recombinant DNA molecule comprising at least a part of that gene, an RNA molecule derived therefrom and a new polypeptide (or protein), as well as a host cell transformed at least with such recombinant DNA molecule, in particular a transformed filamentous fungus. In addition, the invention relates to processes for the use of the new protein or the new host cell in enzymatic conversions using monooxygenases, more specifically enzymes from the cytochrome P450 superfamily.

In particular, the invention relates to the isolation, characterization and use of a gene coding for an NADPH cytochrome P450 oxidoreductase from filamentous fungi. Further, this application discloses the use of this gene for increasing cytochrome P450-mediated enzymatic activities.

In nature, monooxygenase reactions are responsible for a large number of conversions of both endogenous and exogenous compounds. The monooxygenases can be divided into many different classes of enzymes. One of those classes is the family of cytochrome P450 monooxygenases. These are enzymes occurring in both prokaryotes and eukaryotes.

The spatial structure of these proteins has been highly conserved during evolution. The most characteristic part of each cytochrome P450 is the presence of a haem group which is covalently bound to the protein by a sulfur-iron bond. The sulfur always originates from a cysteine molecule. The iron atom is located in the middle of a porphyrin ring with which the atom has four bonds. The sixth ligand of the iron atom is involved in the catalysis of the reaction. An oxygen is bound to it during the process. Accordingly, the manner in which the haem group is bound to the protein is characteristic of this class of proteins and provides for a specific absorption peak at approximately P450 nanometer (CO reduced form).

Cytochrome P450 proteins are only functional in an enzyme complex together with one or two other proteins which take care of the transfer of electrons from NAP(P)H to the active center of the cytochrome P450.

Roughly, the class of cytochrome P450 enzyme systems can be divided in two subclasses, viz. the eukaryotic microsomal P450s and the class of "prokaryotic-like" P450s. The general characteristics of the eukaryotic microsomal P450s are that they are bound to the membranes of eukaryotes, that they are two component enzyme systems (reaction specific cytochrome P450 and general NADPH cytochrome P450 oxidoreductase) and that the reaction is NADPH-dependent. The general characteristic of the other subclass is that the enzyme complexes consist of three components. This subclass can be further subdivided in two groups, viz. the bacterial P450s (soluble, mostly NADH dependent, complex consisting of a reaction specific cytochrome P450 and a general Fe-S protein and NADH reductase) and the eukaryotic P450s occurring in organelles such as mitochondria (membrane-bound, mostly NADPH dependent, complex consisting of reaction specific cytochrome P450 and a general Fe-S protein and NAD(P)H reductase).

The cytochrome P450 proteins of the first subclass accordingly complex to the membrane with the protein NADPH cytochrome P450 oxidoreductase. This protein, like cytochrome P450s, is found in the microsomal

(endoplasmatic reticulum) membrane of lower eukaryotes, plants and animals.

Less is known about the structure of this protein than about that of cytochrome P450s, but here, too, a high degree of structural conservation must have occurred during the evolution, in view of the fact that functional exchange of NADPH cytochrome P450 oxidoreductase has been demonstrated between yeast and mammal systems.

In microsomal enzyme systems of various eukaryotes, it has been demonstrated in different cases that particularly after induction of a specific cytochrome P450, the oxidoreductase could be present in a minor proportion. Thus, the amount of reductase would be determinative or the rate of conversion of compounds to be modified by cytochrome P450. Increasing the amount of NADPH cytochrome P450 oxidoreductase in the cell could then increase the biocatalytic activity (in respect of reactions catalyzed by P450s).

To date, little is known about cytochrome P450 enzyme systems in filamentous fungi. Only a few cytochrome P450 enzymes have been characterized and rather little knowledge is available about the way of electron transfer from NADPH to the cytochrome P450. It has been demonstrated that with fungi, too, the transfer of electrons occurs most probably through a NADPH cytochrome P450 oxidoreductase. In this connection, use was made of immunological and biochemical experiments (Scala, Matthews, Costa & Van Etten (1988) Experimental Mycology 12, 377–385). They show in this article that NADPH cytochrome P450 oxidoreductases of other filamentous fungi can complement the endogenous NADPH cytochrome P450 oxidoreductase in reconstitution experiments with pisatin demethylase from *Nectria haematococca* (CYP 57) using an in vitro pisatin demethylase assay.

These authors further show an immunological relationship between the NADPH cytochrome P450 oxidoreductases of filamentous ascomycetes tested by them, a relationship which is absent with other NADPH cytochrome P450 oxidoreductases, such as those of the yeast *Saccharomyces cerevisiae*.

To date, a limited number of genes coding for cytochrome P450 enzymes from filamentous fungi have been cloned. Further, some biochemical/biotransformation knowledge about a few cytochrome P450 enzymes is available. The genes which have been cloned to date code for a lanosterol 14α-demethylase from *Penicillium italicum* (CYP51), benzoic acid para-hydroxylase from *Aspergillus niger* (CYP53), a cycloheximide inducible cytochrome P450 from *Neurospora crassa* (CYP54), a nitrate/nitrite inducible cytochrome P450 from *Fasarium oxysporum* (CYP55) and pisatin demethylase from *Nectria haematococca* (CYP57). In particular with regard to the first (CYP51) and the last (CYP57) enzyme system, a fair amount of biochemical knowledge is available. The CYP51 enzyme is the point of application for very many fungicides and has therefore been studied from the point of view of use for years. In addition, biochemical knowledge has been published with regard to hydroxylation of biphenyl-like compounds by Aspergillus species.

Heterologous expression of a fungal cytochrome P450 in a non-related fungal host has been described for CYP57. It has been found that the mere transfer of the gene coding for pisatin demethylase from *N. haematococca* to *Aspergillus nidulans* leads to the new host acquiring the ability to demethylate pisatin. This indicates that all other components required for the formation of a functional enzyme complex in *Aspergillus nidulans* were already present and that they

are also capable of forming in vivo an active enzyme complex with a heterologous cytochrome P450.

Increasing the production of a specific cytochrome P450 enzyme (for instance through the introduction of additional gene copies coding for this enzyme) definitely does not always ensure an increase of the total activity of this specific enzymatic reaction. In fact, the introduction of several P450 enzymes to increase the enzymatic activity is suitable only to a certain extent. This will appear from a publication (Van Gorcom et al. (1990) Mol. Gen. Genet. 223,192–197) describing the cloning and characterization of the gene coding for the cytochrome P450 enzyme benzoic acid para-hydroxylase from *Aspergillus niger*. This publication also describes an attempt to improve the benzoic acid para-hydroxylase activity of *A. niger*. This was attempted by increasing the production of the cytochrome P450 in question. However, the increase of the production of the benzoic acid para-hydroxylase did not have a positive effect on the benzoic acid para-hydroxylase activity. Apparently, another component of the enzyme complex was limiting. It could be that the P450 enzyme is not supplied with sufficient electrons which are necessary to enable the conversions to be carried out.

As yet, little is known with regard to the manner in which these cytochrome P450 enzymes are supplied with their electrons. The donor in the microsomal systems described here is NADPH, but it has not yet been demonstrated in detail how the transport proceeds. It has been demonstrated biochemically that, as with other eukaryotic microsomal systems, this is probably done by one enzyme: the NADPH cytochrome P450 oxidoreductase. It is proposed that the oxidoreductase could be that limiting factor.

Accordingly, although a number of P450 enzymes for particular filamentous fungi have been described, the corresponding oxidoreductases from those specific fungi were not available up to the time of the present invention.

SUMMARY OF THE INVENTION

According to the invention, the enzymatic activity of P450 enzyme systems in filamentous fungi can be improved by means of a new gene coding for a new oxidoreductase, by means of new oxidoreductase and/or by means of the new microorganisms hereinafter described in detail.

Whenever in the present specification 'polypeptide' or 'protein' is mentioned, these terms are understood to refer inter alia to a polypeptide which is at least partly coded for by a DNA sequence according to the invention. Examples include fragments and/or derivatives of the new oxidoreductase which exhibit the desired activity.

The only thing that could possibly be used up to the time of the invention to improve cytochrome P450 reactions in fungi are cytochrome P450 oxidoreductases from other, non-related organisms such as yeasts, plants, insects and mammals. These are the organisms whose corresponding genes are available. In the endeavor to use homologous systems which are adapted to the host as best as possible, the use of a cytochrome P450 oxidoreductase stemming from fungi is naturally to be preferred.

On the basis of the known P450 cytochrome oxidoreductases from other organisms, it should be possible with standard techniques to pick up, isolate and clone the oxidoreductase from filamentous fungi.

Surprisingly, the current techniques, such as nucleic acid hybridizations and standard PCR (Polymerase Chain Reaction), failed in the present case. As will be described in the examples with the present invention, it has proved to be impossible to find a gene coding for the oxidoreductase of filamentous fungi with probes (single-stranded DNA-pieces which are complementary to conserved sequences in other oxidoreductases) which have been prepared in the conventional manner.

In addition, it has been found that the enzyme normally used for PCR (Taq-DNA-polymerase) is also unsuitable for picking up the gene that codes for an oxidoreductase of a filamentous fungus using the primers selected by us, which are based on conserved domains in other NADPH cytochrome P450 oxidoreductases.

Only through the choice of a very highly degenerated probe in combination with an unconventional polymerase have we succeeded in obtaining the gene according to the invention.

The new gene comprises a sequence as shown in FIG. 1 (SEQ. ID NO:1). The amino acid sequence derived therefrom is shown in FIG. 2 (SEQ ID NO:2).

For different kinds of uses it is interesting to increase the biocatalytic activity of a fungal cell. The enzymatic activity level naturally present is often not high enough for more industrial uses of these enzyme systems, for example the detoxification of chemicals and toxins in waste, soil, water and air (environmental uses), the modification and detoxification of (raw materials for the production of) foodstuffs and cattle feeds and the use of fungi, or enzyme complexes isolated therefrom, for carrying out one or more enzymatic conversions for the purpose of the synthesis of a chemical or pharmaceutical intermediate or end product. It is also conceivable that a particular product can be produced entirely through fermentation of a fungus. If this involves one or more cytochrome P450 enzymes, the production can be improved by a strain with an increased NADPH cytochrome P450 oxidoreductase activity.

Cytochrome P450 enzyme activities are often the velocity-limiting steps in a pathway. The increase of this type of enzyme activities will therefore have a positive effect on the extent and rate of conversion of compounds synthesized or degraded by such pathway. Examples of conversions (partly) catalyzed by cytochrome P450 are the para-hydroxylation of benzoic acid and other aromatics, various hydroxylations of biphenyl-like compounds, the specific hydroxylation, epoxidation and demethylation of aromatic and non-aromatic compounds with a multiring structure (such as PAKs and steroid-like molecules), the terminal hydroxylation of (short and long chain) alkanes and fatty acids, the demethylation of phytoalexins, etc., etc. At present it is assumed that many cytochrome P450 enzymes are yet to be discovered. However, in the case of these as yet unknown enzyme activities, too, the invention described here can contribute to the increase of the activities in question.

The results of the invention disclosed here are not limited exclusively to the organism with which the investigation has been carried out, i.e. Aspergillus. Two important reasons for this can be mentioned. Firstly, the conservation among homologous genes of filamentous fungi is mostly such that genes isolated from Aspergillus can serve as a tool for (simply) obtaining the corresponding genes from other filamentous fungi. In view of the conservation in the family of cytochrome P450 oxidoreductases, it is expected that this will definitely apply to this gene too. Secondly, the cprA gene (the gene coding for the oxidoreductases according to the invention) of *Aspergillus niger* can also be used directly in other filamentous fungi as a supplier of additional NADPH cytochrome P450 oxidoreductase activity. It has

been found in the past years that the isolation of genes from fungi through heterologous hybridization with the corresponding gene from yeasts as probes is quite often unsuccessful. In these cases the use of a corresponding gene stemming from a filamentous fungus often did prove successful. Even the isolation of genes through PCR experiments using as primers oligonucleotides designed starting from (possibly) conserved regions in yeast genes did not prove by any means to be always a useful approach.

It has further been found that genes from filamentous fungi can be eminently used for carrying out corresponding tasks in other fungi. The promoters are normally functional and introns, too, unlike in the case of yeast, are correctly spliced by other fungi. Exchange of proteins in fungi forming part of a more complex system, which also holds, for instance, for cytochrome P450 enzyme systems, has already been found to be a real possibility several times. Both a subunit of tubulins and a nucleo-coded subunit of a mitochondrial ATPase complex have been found to be exchangeable between different kinds of filamentous fungi in vivo. In vitro, the exchangeability of NADPH cytochrome P450 oxidoreductase between different filamentous fungi has already been demonstrated.

With regard to the activity of the NADPH cytochrome P450 oxidoreductase, it is conceivable that modifications on the protein (provided in directed or non-directed manner) can lead to an NADPH cytochrome P450 oxidoreductase with better properties in general or for a specific use. This kind of adjustments to the protein (via adjustments in the gene) have become possible through the present invention and therefore fall within the scope of protection.

It is also known, in respect of other organisms, that fusion of a specific cytochrome P450 with the NADPH cytochrome P450 oxidoreductase, to be effected through a fusion of the genes in question, can lead to enzymatically active molecules. In a number of cases, this may even have a positive effect on the total enzymatic activity of the cytochrome P450 in question. Such a fusion is also understood to fall within the scope of the present invention.

It is highly doubtful whether it is possible to eliminate the NADPH cytochrome P450 oxidoreductase activity in the cell. This enzyme is involved in a number of activities which are essential to the cell, so that a mutation (directed or randomly provided) in the genome which leads to the elimination of the NADPH cytochrome P450 oxidoreductase activity is probably lethal. However, it is quite plausible that a change of the expression signals of the gene that codes for NADPH cytochrome P450 oxidoreductase (cprA) can also lead to the desired effect (a change (mostly an increase) in the NADPH cytochrome P450 oxidoreductase activity). The data presently available indicate that the promoter of the *Aspergillus niger* cprA gene is not particularly strong. On the basis of the knowledge becoming available through this invention, the strategy which will lead to an increase in the NADPH cytochrome P450 oxidoreductase activity through modification of the promoter and/or the 5' untranslated part of the cprA gene is simple to specify. It can be effected through modification (directed or random) of its own sequences involved in the initiation of mRNA and protein synthesis of the cprA gene. It can also be effected through replacement of the promoter and/or a part of the region of the cprA gene coding for the 5' untranslated part of the mRNA by other, preferably more efficient, sequences (synthetic or stemming from a different gene).

A totally different use of the gene is the use of this gene, or parts thereof such as the promoter, as a reporter for the detection of inducing agents zoals toxins, xenobiotics and the like. Because of the general character of this enzyme it is quite conceivable that in this manner a 'wide-spectrum' diagnostic can be developed. In this connection one could for instance consider fusion of the expression signals of the gene to a simply detectable reporter gene (coding, for instance, for β-galactosidase, β-glucuronidase or luciferase), or detection methods based on immunological techniques.

Although the in vitro use of cytochrome P450 enzyme systems is not yet employed on a large scale, this invention provides the possibility of producing one of the components of such an in vitro enzyme system (the NADPH cytochrome P450 oxidoreductase) on a large scale. This invention also contemplates the making of fungal strains which produce both the cytochrome P450 in question and the NADPH cytochrome P450 oxidoreductase in increased amounts, so that the enzyme complex can be isolated therefrom in simpler manner and in larger amounts.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the DNA sequence of SEQ ID NO:1.

FIG. 2 shows the amino acid sequence of SEQ ID NO:2.

FIG. 3 shows a representation of the picked-up positive λ-clone, λ19-1 and the subclones constructed therefrom, pCPR2 and pCPR7. (S=SalI, E=EccRi, Bg=BglII, K=KpnI, N=NcoI)

FIG. 4 shows a representation of the insert of pCPR2, on which the CPR gene is located. The regions whose sequence has been determined are indicated by arrows. (Bg=BglII, K=KpnI, N=NcoI, X=XhoI, P=PstI)

FIG. 5 is a representation of the 14 dm gene from *Penicillium italicum*.

FIG. 6 is a graph showing fungicide sensitivity as set forth in Table VI.

DETAILED DESCRIPTION OF THE INVENTION

Experimental

Described herein is the cloning and characterization of a gene which codes for a NADPH cytochrome P450 oxidoreductase or a filamentous fungus.

This involves in particular the cprA gene of *Aspergillus niger* ATCC 1015. Initially, an attempt was made to isolate the gene that codes for NADPH cytochrome P450 oxidoreductase from *Aspergillus niger* through heterologous hybridization. The corresponding genes of the yeasts *Saccharomyces cerevisiae* and *Candida tropicalis* were used as probes. These experiments did not lead to positive results (see Example I).

Then an attempt was made to isolate the gene through PCR. On the basis of a comparison of known protein sequences of NADPH cytochrome P450 oxidoreductases from other organisms, primers were designed which code for amino acid sequences from the most conserved regions. These primers were used in PCR experiments as described in Example II. This PCR gave various bands, but not at the expected heights. Subsequently, after a gel electrophoresis of the PCR mixture, the region of the expected size was excised from gel and the (non-visible) DNA was isolated therefrom. Then a renewed PCR reaction was performed on this DNA with the same primers. This PCR did yield a product of the expected size. This PCR product was isolated and cloned into pUC19. However, DNA sequence analysis of the isolated clone yielded a DNA sequence which did not

show any similarity whatever to a known NADPH cytochrome P450 oxidoreductase sequence.

When this attempt proved negative too, an attempt was made to isolate the gene by means of a probe synthesized in an alternative way (Example III). A PCR reaction was performed using *Aspergillus niger* genomic DNA as template and degenerated primers derived from conserved regions from the NADPH cytochrome P450 oxidoreductase. In this experiment, in contrast with the conventional enzyme (Taq DNA polymerase), now a different thermoresistant DNA polymerase was used, viz. the so-called Tth DNA polymerase. This enzyme comes from the bacterium *Thermus thermophylus* and is generally recommended for use in PCR experiments with RNA as template. By means of this enzyme, too, many different bands were found, among others a band of the expected size. This band was isolated and cloned in pUC19. Sequence analysis of a single clone, pCPR1, gave a clearly recognizable similarity (at protein level) to the NADPH cytochrome P450 oxidoreductase sequences known at that time. This clone was used as a probe for the purpose of the isolation of the NADPH cytochrome P450 oxidoreductase gene of *Aspergillus niger* from a gene library.

The *Aspergillus niger* cprA gene was isolated on a 7 kb EcoRI-SalI fragment (pCPR7) and a 3.7 kb BglII-KpnI fragment (Example IV). The DNA sequence of the complete 3.7 kb BglII-KpnI fragment was determined. It is shown in FIG. I (SEQ ID NO: 1). In this DNA sequence an open reading frame was found, coding for 692 amino acids. This open reading frame is interrupted once by a DNA sequence coding for an intron (see Example V). The start of the mRNA was also determined (Example VI).

A comparison of the deduced amino acid sequence with the known amino acid sequences of NADPH cytochrome P450 oxidoreductases from other organisms indicated that the identicality with the NADPH cytochrome P450 oxidoreductase most closely related on the basis of amino acid sequences (that of the yeast *Saccharomyces cerevisiae*) is only 40%. In various "conserved" regions of the *Aspergillus niger* NADPH cytochrome P450 oxidoreductase, too, the identicality is not 100%.

The expression of the *Aspergillus niger* cprA gene on an mRNA level was also analyzed. The expression of the gene is found to be not very high, which provides possibilities for the improvement of the expression by improving the cprA promoter or by replacing the cprA promoter by a different, stronger promoter.

The effect of the introduction of several copies of the *Aspergillus niger* cprA gene on the extent of NADPH cytochrome P450 oxidoreductase activity was studied, as was the effect thereof on the cytochrome P450 activity.

Example XI describes that the introduction of several copies of the cprA gene into *Aspergillus niger* can lead to a strongly increased NADPH cytochrome P450 oxidoreductase activity. The effect of this on the benzoic acid parahydroxylase activity of this strain (wild-type) is also presented in this example.

Example XII describes that the introduction of several copies of the cprA gene into an *Aspergillus niger* transformant which already contains several copies of the benzoic acid parahydroxylase gene (bphA) (and, as a result, has an increased production of the benzoic acid para-hydroxylase) can lead to a very substantial increase of the BPH activity. These experiments clearly show that the NADPH cytochrome P450 oxidoreductase gene described in this patent application, when introduced into a fungus, can have a very positive effect on cytochrome P450 activities.

EXAMPLE I

Cloning of the cprA gene of *Aspergillus niger* by means of hetearologous hybridization

Insofar as they are not described in detail, techniques were carried out as described (Sambrook, Fritsch & Maniatis (1989) Moleculair cloning, Cold Spring Harbor Laboratory Press, U.S.A.).

An attempt was made to isolate the gene coding for cytochrome P450 reductase of *Aspergillus niger* ATCC 1015 by means of heterologous hybridization experiments. To that end, chromosomal DNA (5 μg) of *Aspergillus niger* and of *Saccharomyces cerevisiae* was digested with HindIII and with EcoRI. This DNA was separated by means of electrophoresis on a 0.8% TBE-agarose gel and transferred to a Hybond N membrane (Amersham). In this way several identical blots were made. After fixation of the DNA on the membrane by baking for 2 hours at 80° C. and 2–4 hours of prehybridization, the blot was hybridized with a ^{32}p labeled (Amersham multiprime DNA labeling kit, in accordance with the manufacturer's directions) cytochrome P450 reductase specific probes coming from *Candida tropicalis* and *Saccharomyces cerevisiae*, respectively. As a probe for the *Candida tropicalis* CPR gene, a 820 bp EcoRI fragment containing a part of the cprI and a 3.7 kb SpeI fragment comprising the entire CPR gene were isolated from plasmid pTS1 (Sutter et al., 1990, J. Biol. Chem. 265 (27), 16428–16436). As a *Saccharomyces cerevisiae* CPR specific probe, a 700 bp BamHI fragment containing an internal part of the CPR and a 3.3. kb PvuII fragment comprising the entire CPR gene were isolated from plasmid pTS20 (Sutter & Loper, 1989, Bioch. Biophys. Res. Comm. 160(3), 1257–1266). The fragments were separated by electrophoresis on a 0.8% TBE-agarose gel and purified by means of the Geneclean kit (Bio101) in accordance with the supplier's directions. The blots were hybridized and washed at 56° C. as described (Sambrook, Pritsch & Maniatis (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, U.S.A.). The last washing step was performed at 56° C. with 6×SSC, 0.1% SDS. With the *S. cerevisiae* probe, clear, specific hybridizing bands could be seen in the lanes with *S. cerevisiae* chromosomal DNA. Experiments with the *C. tropicalis* probes were exclusively performed on chromosomal DNA of *Aspergillus niger*. Surprisingly, with neither probe hybridizing bands were found in the lanes with *Aspergillus niger* chromosomal DNA, possibly caused by too limited a homology between the *Aspergillus niger* cytochrome P450 reductase gene on the one hand and the *S. cerevisiae* and *C. tropicalis* cytochrome P450 reductase genes on the other.

EXAMPLE II

Cloning of the cprA gene of *Aspergillus niger* by means of PCR with Taq DNA polymerase

In view of the negative results in the heterologous hybridization experiments, it was decided to attempt to synthesize a cpr specific fragment by means of the PCR technique. Degenerated primers were resigned, based on conserved sequences in known cytochrome P450 oxidoreductase genes of rat (Porter & Kasper (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 973–977); *Candida tropicalis* (Sutter et al. (1990) J. Biol. Chem. 265, 16428–16436); and *Saccharomyces cerevisiae* (Sutter & Loper (1989) Bioch. Biophys. Res. Comm. 160, 1257–1266) (Table I). At the ends, additional sequences were provided, coding for specific restriction sites (5' PCR primers were provided with an EcoRI restriction site, 3' primers with a BamHI restriction size) to facilitate the cloning of PCR products and to increase the specificity of the procedure.

TABLE I

	EcoRI
MBL 997	5'- CCG GAA TTC CA(G/A) ACN GGN ACN GCN GA(G/A) GA - 3'(SEQ ID NO:3)
	degenerated 1024 times, FMN binding site
	EcoRI
MBL 998	5' - CCG GAA TTC GGN GAN CCN ACN GA(C/T) AA(C/T) GC - 3'(SEQ. ID NO:4)
	degenerated 1024 times, FMN binding site
	BamHI
MBL 999	5' - CGC GGA TCC GGN CCN A (C/T)N A(G/T/A) (G/T/A) ATN AC - 3'(SEQ. ID NO:5)
	degenerated 4608 times, NADPH binding site
	BamHI
MBL 1000	5' - GCG GGA TCC T(C/G) (C/T) TGN AC(G/A) TAN AC(C/T) TT - 3'(SEQ ID NO: 6)
	degernated 256 times, NADPH binding site
	BamHI
MBL 1001	5' - CGC GGA TCC GGN CC(G/T/A) ATC AT(G/T/A) ATN AC - 3'(SEQ ID NO:7)
	degenerated 144 times, NADPH binding site

Table I, sequences of the primers used for the isolation of a cpr specific probe. N stands for G/A/T/C.

For the PCR reactions, Taq DNA polymerase (Perkin Elmer) was used. As template, *Aspergillus niger* chromosomal DNA was used. By way of check, PCR reactions were performed using as template 10 ng of the plasmid pTS20 (on which is located the entire cytochrome P450 reductase gene of *Saccharomyces cerevisiae*) and chromosomal DNA of *Saccharomyces cerevisiae*.

Template and primers were denatured for 10 minutes at 94° C., followed by 25 cycli (1 minute 94° C., 1 minute 43° C., 2 minutes 72° C.). After 25 cycli, incubation took place for 5 minutes at 72° C.

When Taq DNA polymerase with plasmid pTS20 as template was used, fragments of the expected size were found when primer combinations MBL997-MBL999, MBL997-MBL1000 and MBL997-MBL1001 were used. The use of primer MBL998 did not in any of the cases lead to synthesis of a product of the expected size.

The use of *S. cerevisiae* chromosomal DNA as template led to a clearly demonstrable product of the expected size when the primer combination MBL997-MBL1001 was used. The use of the combination of MBL997 with the primer MBL999, degenerated 4608 times, led to a weakly demonstrable product of the expected size. In all of these cases a large amount of a specific product of smaller sizes was found. With the primers MBL998 and MBL1000, using *S. cerevisiae* chromosomal DNA as template, no correct PCR product could be demonstrated.

When *Aspergillus niger* chromosomal DNA was used as template, and using Taq DNA polymerase, no product of the expected size was found with any of the primer combinations used. What could be demonstrated in all possible combinations was a large amount of a specific, too small, product.

The total material of a single experiment with the primer combinations MBL997–999, MBL997–1000 and MBL997–1001 was separated through electrophoresis on a 0.8% TBE-agarose gel. A piece of gel at the level of a marker band of 1.2 kb (the expected size) was excised. At this location in the gel no DNA was visible. DNA from this gel fragment was isolated by means of the freeze squeeze method. The so obtained material was used as template for a new PCR reaction. Surprisingly, after this reaction a product of the expected size was found when using the primer combinations MBL998/999 and MBL998/1001. However, with all of the primer combinations used (MBL997–MBL999,1000,1001 and MBL998–MBL999, 1000,1001) mainly smaller, a specific, products were found. The product of the expected size that was found was separated through gel electrophoresis on a 0.8% Low Melting Point agarose gel whereafter DNA was isolated by means of β-agarase (New England Biolabs, in accordance with the directions of the supplier). The DNA obtained was digested with EcoRI and BamHI and cloned into the EcoRI and BamHI sites of pUC19. After transformation to *Escherichia coli*, miniscreen DNA was isolated from 12 ampicillin resistant colonies. After digestion of the obtained miniscreen DNA with EcoRI and BamHI, only one construct was found to contain an insert of the correct size. Of this a large plasmid isolation was done, using a Qiagen Tip500 in accordance with the supplier's instructions. The DNA sequence of this construct was analyzed by dideoxy sequence analysis (Pharmacia T7 sequencing kit, according to the supplier's instructions) using the M13 universal and reverse primer. In the DNA sequence as determined, no homology with known cytochrome P450 reductase genes of other organisms was found.

EXAMPLE III

Cloning of the cprA gene of *Aspergillus niger* by means of PCR with Tth DNA polymerase

As an alternative, it was decided to perform PCR experiments using, instead of 1 unit Taq DNA polymerase, 0.1 unit

Tth DNA polymerase (Spaero Q), an enzyme which is particularly suitable for RT-PCR (reverse transcriptase PCR) experiments. Using plasmid pTS20 as template, this resulted in products of the expected size when using primer combinations MBL997–999, MBL997–1000 and MBL997–1001. When *Aspergillus niger* chromosomal DNA was used as template, this resulted in a large amount of products when using primer combination MBL997–1001. The majority of these products were found to be smaller than the expected size. Surprisingly, when using these degenerated primers in combination with Tth DNA Polymerase, in addition to the a specific products, a product of the expected size was found. The PCR products were separated by means of gel electrophoresis on a 1% TBE-low melting point agarose gel, whereafter the presumably correct product was isolated by means of β-agarase.

After digestion of the obtained DNA with EcoRI and BamHI, the fragment was cloned into the EcoRI and BamHI restriction sites of plasmid pUC19. After transformation to *E. coli*, miniscreen DNA was isolated from 12 ampicillin resistant colonies. After digestion of the miniscreen DNA with EcoRI and BamHI, 4 of the 12 preparations were found to contain a plasmid with an insert of the correct size.

Of these four preparations, of one construct (pCPR1) a large DNA isolation was done by means of a Qiagen Tip500. Of this DNA, the sequence of a part of the insert was determined by means of the M13 universal and reversed primer. This sequence was found to contain regions which, at an amino acid level, were clearly homologous to other cytochrome P450 reductase genes.

EXAMPLE IV

Screening of *Aspergillus niger* λ-gene library with the cpr specific probe

Genomic DNA of the *Aspergillus niger* strain ATCC 1015 was partially digested with Sau3AI and separated by gel electrophoresis. DNA of between 13 and 17 kb in size was isolated from the gel and the DNA was isolated. The isolated DNA was cloned into vector λEMBL3, digested with BamHI. This genomic library was screened with the cpr specific probe isolated from plasmid pCPR1 (the 1.2 kb EcoRI-BamHI fragment, see FIG. 4). The isolated EcoRI-BamHI restriction fragment, with the cpr specific PCR fragment, was radioactively labeled with ^{32}P-dCTP using the multiprime DNA labeling kit. (Amersham) in accordance with the supplier's instructions. Labeling reactions were performed for 3 hours at room temperature. The probe was purified by spin column filtration over a 1 ml Sephadex G50-medium column. About 40,000 pfu were screened (about 20 times the genome). In the first screening 13 positive plaques were isolated. These were screened again. After the second screening, 3 positive plaques remained (λ5-3, λ11-1 and λ19-1). Through Southern analysis it was demonstrated that all clones contained fragments which hybridized with the cpr specific probe. The restriction patterns of these clones were clearly related but not identical. A restriction map of clone λ19-1 was made (see FIG. 3). A 7 kb EcoRI-SalI fragment and a 3.7 kb BglII-KpnI fragment of clone λ19-1, which contain the coding region of the cprA gene, were isolated and cloned into vector pBluescript-SK II$^+$ (Stratagene) digested, respectively, with EcoRI and SalI and with BamHI and KpnI. This resulted in the vectors pCPR7 and pCPR2 (see FIGS. 3 and 4).

EXAMPLE V

Sequence analysis of the *Aspergillus niger* cprA gene

The sequence of both strands of the *Aspergillus niger* BglII-KpnI fragment cloned into plasmid pPCPR2 was determined according to the "dideoxy chain terminating" method (Sanger et al. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467) using the ^{35}S-dATP T7 DNA sequencing kit (Pharmacia) in accordance with the supplier's instructions. Sequence analysis of subclones was performed using the M13 universal and reverse primers. Parts of the sequence were determined using synthetic oligonucleotides, using pCPR2 ds-DNA as template (see FIG. 4). Sequence information was analyzed with GCG software (Devereux et al., (1984) Nucleic Acids Res 12, 387–395).

In the DNA sequence a region possibly coding for an intron was found (FIG. 2, position 2367 . . . 2437). The actual absence of this intron in the mRNA of the cprA gene was demonstrated with RT-PCR experiments. In them the expected difference in size of the PCR product was found when using primers located around the intron, with chromosomal DNA or isolated total RNA as template (GeneAmp RNA-PCR kit, Perkin Elmer).

The deduced amino acid sequence of the *Aspergillus niger* cprA gene was found to show a 39–46% identicality to the amino acid sequences of the CPRs of the yeasts *Saccharomyces cerevisiae, Candida tropicalis* and *Schizosaccharomyces pombe* and the genes from rodents and an approximate 35% identicality to the CPR amino acid sequences of plants published to date.

EXAMPLE VI

Determination of transcription start

For determining the transcription start of the *Aspergillus niger* cytochrome P450 reductase gene, a primer extension experiment was performed. To that end, two primers, PE50 and PE100, were designed. These primers were based on sequences located on, respectively, 50 bp and 100 bp 3' of the ATG and read in the direction of the ATG.

PE50

5' CCA-CGC-TAC-CCA-C 3 (SEQ ID NO:8)

PE100

5' GAG-GCA-TAC-GGG-TC 3' (SEQ ID NO:9)

RNA was isolated from *Aspergillus niger* transformant T16 (van Gorcom et al., (1990) Mol Gen Genet 223, 192–197). Mycelium was cultured and ground in liquid nitrogen. The fine powder was resuspended in 1 ml RNAzol (Cinna/Biotecx) whereafter the isolation was continued in accordance with the supplier's instructions.

Of both primers 100 ng was kinated with 5 μl ^{32}P-γATP. Reactions were performed for 30 minutes at 37° C. as described (Sambrook, Fritsch & Maniatis (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, U.S.A.). After the incubation the primers were purified by filtration over a Sephadex G50 spin column. Added to this primer was NH_4 acetate to a final concentration of 2M and 2.5 volume ice-cold Ethanol. Primers were precipitated at –20° C. for 60 minutes. Primers were pelletized by centrifugation at 4° C., 14000 rpm. for 30 minutes, whereafter the pellets were washed with 70% ethanol and were resuspended in 20 μl TE.

Primers were ex-ended with M-MLV reverse transcriptase for 1 hour at 37° C. The reaction mix (25 μl) contained 10 μg RNA, 50 mM NaCl, 34 mM Tris-HCl/pH 8.3, 6 mM $MgCl_2$, 5 mM DTT, 200 U M-MLV reverse transcriptase (Gibco), 0.5 mM dGTP, 0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dTTP and 5 μl (25 ng) of the kinated primer.

The reactions were stopped by the addition of 2 μl 0.5M EDTA, followed by an alcohol precipitation (60 minutes –20° C.). After centrifugation (30 minutes, 14000 rpm, 4° C.) the pellet was resuspended in 5 μl distilled water. To the mix was added 3.4 μl stop mix (Pharmacia T7 polymerase sequencing kit). Before being applied to gel, the samples were denatured by incubation at 95° C. for 5 minutes,

whereafter they were directly placed on ice. One half of the sample was applied to a 8% wedge shape denaturing polyacrylamid gel.

By way of check, sequence reactions were performed with both primers (Pharmacia T7 polymerase sequencing kit) with plasmid pCPR2 as template and applied to gel simultaneously with the primer extension mix. After 2 hours of electrophoresis (1100 V) the gel was dried and an X-ray sensitive film was exposed.

Strikingly, results with primer PE50 were much stronger than with primer PE100. In particular in the lanes with primer PE50 a great many bands were visible. Combination of both results led to the designation of three possible transcription start positions, viz. at 86 bp (ctcActc), 55 bp (cagActcg) and 37 bp (aagTcgg) before the ATG. The last band was not found when primer PE100 was used.

EXAMPLE VII

Transformation of *Aspergillus niger* strains

The *Aspergillus niger* strains N204 (ATCC 1015, csp, met, Boschloo et al., Appl Microbiol Biotechnol (1990) 34, 225–228; contains 1 copy of the bphA gene) and T18 (*Aspergillus niger* N271 provided with about 12 copies of the bphA gene; van Gorcom et al. (1990) Mol Gen Genet 223, 192–197) were cultured in 250 ml complete medium (minimal medium supplemented with 0.1% Cas amino acids, 0.5% Yeast extract, 0.1 mg/ml methionine, 1 μg/ml pyridoxine) in 21 Erlenmeyer flasks. The cultures were inoculated with 1.10^6 spores per ml and then incubated for 18 hours in an air-agitated incubator at 35° C. and 300 rpm. After incubation, mycelium was harvested by filtration through a miracloth filter (Calbiochem). This mycelium was then protoplasted and the protoplasts isolated therefrom were transformed as described by Yelton et al., Proc. Natl. Aca. Sci. USA 81 (1984) 1470–1474.

Samples of 100 μl protoplasts were transformed with, in all, 1 μg circular pAN7-1 DNA, 9 μg circular pCPR2 DNA, and 2 μl of a 1M ATA (Aureen tri-carboxyl acid) solution. Control tubes contained no DNA or only 1 μg pAN7-1 (Punt et al., Gene 56 (1987) 117–124). The transformed protoplasts were then plated on minimal medium agar plates supplemented with 1.2M sorbitol, 1 μg/ml pyridoxine, 0.1 mg/ml methionine and 100 μg/ml hygromycin. The plates were incubated for 10 days at 35° C. After 4 days the first sporulating transformants became visible. Transformants were twice applied with a brush to form pure cultures on minimal medium agar plates with 100 μg/ml hygromycin, 0.1 mg/ml methionine and 1 μg/ml pyridoxine.

Per transformation 10 to 40 hygromycin resistant transformants could be formed. On transformation plates of protoplasts which had been treated without DNA, no hygromycin resistant transformants could be found after 14 days of incubation at 35° C.

EXAMPLE VIII

Screening of transformants

T18 multiple copy cprA transformants

The transformants from Example VII were tested for their hygromycin resistance level by plating of spores on minimal medium agar plates supplemented with 500 μg/ml hygromycin, 0.1 mg/ml methionine and 1 μg/ml pyridoxine.

For selecting transformants with an increased cprA mRNA level, an RNA colony hybridization experiment was performed according to a modified version of the protocol as was described for *Saccharomyces cerevisiae* by Stepien and Butow (1992, Nucl.Acids Res. 18(2) p380). Minimal medium agar plates were inoculated with spores of transformants, covered with a Hybond-N filter, and subsequently incubated at 25° C. until mycelium was just visible but spore formation had not developed yet. The filters were transferred to a 500 μl drop of sorbitol buffer (1.2M sorbitol, 0.1M sodium citrate/pH 5.8, 0.1M EDTA, 50 mM μ-mercapto-ethanol), incubated for 5 minutes and subsequently transferred to a sheet of Whatmann 3 MM filtering paper. After drying for 5 minutes on the Whatmann paper, the filters were transferred to a Petri dish with a drop (500 μl) of sorbitol buffer to which 10 mg/ml Novozym 234 (NOVO Nordisk) had been added. The Petri dish was closed airtightly with parafilm and the mycelium was protoplasted for 1 hour by incubation at 35° C. After protoplasting, the protoplasts were lysed by transferring the filter to a 500 μl drop of lysis buffer (2% SDS, 7.3% formaldehyde, 50 mM Tris-HCl/pH 7.5, 10 mM EDTA) followed by an incubation of 5 minutes at room temperature. The RNA was blotted on the filter by transferring the filter, with lysed protoplasts, to a sheet of Whatmann 3 MM filtering paper until the filter was dry (about 5 minutes). This blotting step was repeated a single time. Finally, the filters were transferred for 1 minute to a drop of 800 μl 6×SSC (Sambrook et al, ***), 0.1% SDS and subsequently dried by transferring the filter to a sheet of Whatmann 3 MM paper. The RNA was fixed on the Hybond N filter through UV crosslinking. To that end, the filter was set on a UV light box for 3 minutes. The filters were hybridized overnight at 65° C. with a ^{32}p-dCTP labeled cprA probe. The filters were washed at 65° C., utilizing 0.2×SSC, 0.1% SDS in the last washing step. After overnight exposure at −70° C., positive transformants could be identified.

N204 multiple copy cprA transformants

Hygromycin resistant transformants of N204 were screened with an RNA colony hybridization in which a cprA specific probe was used. A second screening was performed with a cytochrome P450 reductase specific filter activity assay. For this purpose, spores of hygromycin resistant transformants were inoculated on minimal medium agar plates supplemented with 0.1 mg/ml methionine and 1 μg/ml pyridoxine. The plates were covered with a Hybond N filter and incubated at 25° C. until mycelium was just visible but virtually no spores had been formed yet. The filters were removed, the mycelium was lysed by freezing in liquid nitrogen and subsequent defrosting. The filter was incubated in the dark for about one hour in 25 ml neotetrazolium solution (5 mg Neotetrazolium, 5 mg NADPH per 25 ml). At locations with cytochrome P450 reductase activity, a clearly pink precipitate is formed.

On the basis of the two screening methods, transformants W13 and W35 were selected as presumable multiple copy cprA transformants.

EXAMPLE IX

DNA analysis of transformants

Mycelium of transformants selected in Example VIII was cultured in 50 ml minimal medium, supplemented with 1 μg/ml pyridoxine, 0.1 mg/ml methionine and 0.1% CAS amino acids. Erlenmeyer flasks (300 ml) were inoculated with 1.10^8 spores and placed in an air-agitated incubator (35° C., 300 rpm). The mycelium was harvested by filtration over miracloth filter (Calbiochem) and washed with 25 ml 0.9% NaCl. The mycelium was immediately frozen in liquid nitrogen. From this mycelium, chromosomal DNA was isolated according to the method described by Kolar et al. (Gene 62 (1988) 127–134). The pellet of the last alcohol precipitation was incorporated in 100 μl distilled water.

Chromosomal DNA (40 μl) was digested with 20 U EcoRI and 20 U KpnI for 6 hours at 37° C. After the digestion equal amounts of DNA were separated on a 0.8% TBE-agarose gel

by means of electrophoresis (18 hours, 35 volts). After separation the DNA was transferred to a sheet of Hybond N filter (Amersham) in accordance with the instructions of the supplier. After fixation of the DNA (2 hours, 80° C.) the blot was prehybridized at 65° C. for 4 hours and subsequently hybridized at 65° C. with a ^{32}p labeled cprA specific probe. The blot was washed at 65° C., the last washing step involving 0.2×SSC, 0.1% SDS. An X-ray sensitive film was exposed at −70° C. For 24 hours.

Seven selected transformants were analyzed in this manner. In one transformant, only the wild-type band was found, in four transformants integration of 1–2 copies of the cprA gene was found and in two transformants integration of several copies of the cprA was found. Of these multiple copy integrants, transformant T18 #5 was selected.

EXAMPLE X

NADPH: Cytochrome P450 (cytochrome c) reductase activity

CPR activity was determined by measuring the possibility of cell free extracts of transformants to reduce cytochrome c. (Adapted from Madyastha et al. (1976) Biochemistry 15, 1097–1102).

Mycelium was cultured for 13 hours in 50 ml minimal medium, supplemented with 0.1% CAS amino acids, 0.1 mg/ml methionine and 0.1 μg/ml pyridoxine. The medium was inoculated with 1.10^6 spores per ml in 300 ml Erlenmeyer flasks in an air-agitated incubator (35° C., 300 rpm). The mycelium was harvested by filtration over miracloth filter (Calbiochem) and washed with 25 ml 0.9% NaCl. Excess buffer was removed by blotting the filter with mycelium between tissues. The mycelium was frozen in liquid nitrogen and ground in a mortar. The fine powder was transferred to an Eppendorf reaction vessel filled with 1 ml ice-cold CPR-extraction buffer (50 mm sodium phosphate buffer/pH 7.8, 20% glycerol, 1 mM EDTA, 1 mM DTT, 5 μg/ml Leupeptin, 4 mM PMSF, 0.2% sodium deoxycholate). The vessels were directly mixed and placed on ice. The extract was incubated on ice for 15 minutes (solubilization step) and centrifuged for 15 minutes at 4° C., 1000 g. The supernatant was transferred to a new vessel and preserves an ice. Cytochrome c reductase activity was measured directly.

To 1 ml CPR-assay mix (0.3M K-phoschate buffer/pH 7.7, 0.1 mM EDTA, 1 mM KCN) was added 2 μl 20 mM cytochrome C. At room temperature (20°–25° C.) 5–20 μl cell free extract was added and incubated for 1–2 minutes. Reactions were initiated by addition of 1 μl 100 mM NADPH. Reduction of cytochrome c was monitored spectrophotometrically for 3 minutes (550 nm).

Protein concentrations were determined with the Bio Rad protein assay kit in accordance with the instructions of the supplier, using BSA as standard.

Table II, Cytochrome C reductase activities (e=21 $mM.cm^{-1}$) of different strains/transformants.

Values coming from two independent cultures are specified. Percentages are based on the average of the duplo experiments. One unit is the amount of cytochrome C (mM) which is reduced per minute per mg total protein.

Strain/ transformant	Estimated cprA copy number	Units	Percentage
N204	1	2.14	100
W13	≧6	20.56	960
W35	≧6	8.12	379
T18	1	2.50	117
T18 #5	≧10	54.57	2.550

EXAMPLE XI

BPH activity in vitro assay

For measuring BPH activity in vitro a not yet optimized assay was used, based on the benzoate dependent consumption of NADPH by the NADPH cytochrome P450 oxidoreductase.

The BPH activity was measured in vitro in *Aspergillus niger* strains T18 (1 copy cprA, 12 copies bphA), strain T18 #5 (6 copies cprA, 12 copies bphA), strain N204 (1 copy cprA, 1 copy bphA) and strain W35 (1 copy bphA, multiple copy cprA)

Mycelium was cultured for 18 hours in 250 ml minimal medium, supplemented with 0.1% CAS amino acids, 0.1 mg/ml methionine and 0.1 μg/ml pyridoxine. The medium was inoculated with 1.10^6 spores per ml in 2 1 Erlenmeyer flasks in an air-agitated incubator (35° C., 300 rpm). The mycelium was harvested by filtration over miracloth filter (Calbiochem) and washed with 250 ml 0.9% NaCl. The mycelium was transferred to 2 1 Erlenmeyer flasks filled with 250 ml induction medium (minimal medium with pyridoxine and methionine in which, instead of glucose, 0.1% benzoate was present as C-source). Mycelium was harvested by filtration over miracloth filtration cloth. After washing with 0.9% NaCl, excess buffer was removed by blotting the filter with mycelium between tissues. The mycelium was frozen in liquid nitrogen and ground in a mortar. The fine powder was transferred to an Eppendorf reaction vessel filled with 1 ml ice-cold CPR-extraction buffer (50 mM sodium phosphate buffer/pH 7.8, 20% glycerol, 1 mM EDTA, 1 mM DTT, 5 μg/ml Leupeptin, 4 mM PMSF, 0.2% sodium deoxocholate). The vessels were directly mixed and placed on ice. The extract was incubated on ice for 15 minutes (solubilization) and then centrifuged for 15 minutes at 4° C., 3500 rpm. The supernatant was transferred to a new vessel and preserved on ice.

BPH activity was measured by spectrophotometrically monitoring (340 nm) the BPH specific consumption of NADPH by the cytochrome P450 reductase. Cell free extract (10 μl) was added to 500 μl BPH assay buffer (100 mM Tris/pH 7.8, 10 nM $MgCl_2$, 200 μM NADPH). The non-specific NADPH consumption was measured for 2 minutes (δ-BA). Benzoate was added (20 μl of a 20 mM solution) and the NADPH consumption was measured for 4 minutes (δ-BA). BPH specific MADPH consumption was determined by the following calculation method:

$$\frac{(\delta + BA)}{\text{min.}} - \frac{(\delta - BA)}{\text{min.}} = \delta BA$$

Units (1 unit=1 μmol NADPH consumed per minute per mg total protein) were calculated by multiplying δBA with the extinction coefficient ($e=6.22.10^{-3}M^{-1}cm^{-1}$).

Protein concentrations were determined with the Bio Rad protein assay kit in accordance with the instructions of the supplier, and BSA was used as standard.

Table IV, in vitro BPH activity. One unit corresponds with the benzoate dependent consumption of NADPH (μM) per minute per mg total protein.

Strain/ Transformant	Estimated bphA copy number	Estimated cprA copy number	Units	Percentage
N204	1	1	365	100
W13	1	≧6	577	158
T18	12	1	899	246
T18 #5	12	≧10	1590	435

EXAMPLE XII

Bph activity, in vivo HPLC assay

Mycelium was cultured for 18 hours in 500 ml minimal medium, supplemented with 0.1% CAS amino acids, 0.1 mg/ml methionine and 0.1 μg/ml pyridoxine. The medium was inoculated with 1.10^6 spores per ml in 2 1 Erlenmeyer flasks in an air-agitated incubator (35° C., 300 rpm). The mycelium was harvested by filtration over miracloth filter (Calbiochem) and washed with 25 ml 0.9% NaCl. Mycelium was subcultured in induction medium (minimal medium supplemented with 0.1 mg/ml methionine and 0.1 μg/ml pyridoxine, while adding, instead of 1% glucose, 0.1% benzoate as C-source). Samples of the medium were taken after 5 hours.

Samples of 2 μl were analyzed by means of HPLC chromatography an a reversed phase C-18 column (Superchem LC 18-DB) at 30° C., using 10 mM sodium citrate buffer/pH 3, 60% methanol as elution buffer. A flow rate of 1 ml/min. was used. Both benzoate and 4-hydroxybenzoate were detected at 245 nm. Reference samples contained 10 μl 1 mM benzoate and 10 μl 1 mM 4-hydroxybenzoate.

Table IV, concentrations of benzoate and 4-hydroxybenzoate after incubation of transformants for 5 hrs in induction medium. Detection by means of HPLC. *Aspergillus niger* can use 4-hydroxy-benzoate as C-source and so metabolize further. As a result, the sums of the last two columns of the amounts of benzoate and 4-OH-benzoate are not identical.

Strain/ Transformant	Estimated bphA copy number	Estimated cprA copy number	Benzoate (mM)	4-OH-Benzoate (mM)
medium	—	—	6.70	0.00
N 204	1	1	1.15	3.20
W35	1	≧6	0.60	3.63
T18	12	1	2.40	3.30
T18 #5	12	≧10	0.0056	3.28

EXAMPLE XIII

To verify the broad action of the cytochrome P450 oxidoreductase gene, experiments were performed in which the effect of CPR overproduction on the activity of a different (in this example even heterologous) cytochrome P450 enzyme was tested.

For that purpose, *Aspergillus niger* strains were constructed which were provided with several copies of the cprA gene of *Aspergillus niger* together with several copies of the gene coding for lanosterol 14α-demethylase (14 dm) from the filamentous fungus *Penicillium italicum*.

Construction of plasmids

For the introduction of the two genes, two plasmids were constructed. Plasmid pCPR2-amdS was constructed by digesting plasmid pCPR2 with NotI and providing it with a functional copy of the amdS gene of *Aspergillus nidulans* (Hynes et al. Mol. Cell. Biol. 3 (1983) 1430–1439), located on a NotI fragment approximately 5 kb in length (the original ends of the chromosomal amdS fragment (EcoRI and SalI) had been replaced with NotI sites). The amdS gene was used as selection marker in transformation experiments. Plasmid p14 dm was constructed as follows. The yeast-expression plasmid YEP24 was digested with BamHI and SalI. Between these sites was ligated an approximately 2.1 kb chromosomal (partial) BamHI-(partial) SalI fragment, on which the entire 14 dm gene from *Penicillium italicum* was located (see FIG. 5).

Transformation of *Aspergillus niger* pCPR2-amdS

A. niger N402 (Bos, PhD Thesis Agricultural University Wageningen NL, 1986) was used as starting strain. Transformation experiments were carried out as described in (Kelly and Hynes EMBO J. 4 (1985) 475–479). *A. niger* N402 was transformed with pCPR2-amdS. Transformants, after being applied with a brush to form pure cultures on selection plates (minimal medium plates with 15 mM CsCl with Acetamide (10 mM) functioning as sole N-source), were further selected for their possibility to use acrylamid as sole N-source. Strains with a high amdS copy number grow better on acrylamid than do strains with a lower copy number (Verdoes et al., Transgenic Research 2 (1993) 84–92).

A selected number of transformants were further analyzed by means of Southern blot analysis. Eventually, transformant AB2-2 (≧10 copies cprA) was selected for further experiments. p14 dm

Plasmid p14 dm was introduced into *Aspergillus niger* by means of cotransformation with plasmid pAN7-1 on which the hph gene is located, which affords resistance to Hygromycin (Punt et al. Gene 56 (1987) 117–124). Transformants were selected on plates with 100 μg/ml hygromycin. Positive transformants were thereafter selected for their resistance to higher concentrations of hygromycin. Transformants that grow well at higher concentrations of hygromycin too, were further analyzed by means of Southern Blot analysis. Eventually transformant AB-D1 (≧10 copies 14 dm) was selected for further experiments.

In transformant AB-D1, in a second transformation experiment, an attempt was made to introduce, in addition to extra 14 dm copies, extra cprA copies. Transformant AB-D1 was for this purpose transformed with plasmid pCPR2-amdS. The further selection procedure was identical to the procedure followed for the isolation of transformant AB2-2. Eventually transformant ABD1.15 (≧10 copies 14 dm, ≧10 copies cprA) was selected for further experiments.

Cytochrome P450 reductase activity

Strain/ Transformant	Estimated 14DM copy number	Estimated cprA copy number	Units
N402	"1"	1	1.05
AB2-2	"1"	≧10	14.6
AB-D1	"1" + ≧10	1	1.65
AB-D1.15	"1" + ≧10	≧10	41.7

Table V. CPR activity in selected transformants, measured as NADPH dependent cytochrome C reduction and expressed in Units (mmol cytochrome c reduced per minute per mg. total protein). Estimated 14 DM copy number: "1" wildtype *A. niger* 14 DM gene, +: extra copies of the *P. italicum* 14 DM gene. Lanosterol 14α-demethylase activity.

Lanosterol 14α-demethylase activity in the four different strains was determined in an experiment wherein the radial growth of the different strains in media with different

concentrations of inhibitor of the lanosterol 14α-demethylase is measured as a measure for the enzymatic activity. For that purpose, mycelial plugs were provided on plates with increasing concentrations of known lanosterol 14α-demethylase inhibitors (DMIs). At a higher concentration of the DMI, an increasingly smaller outgrowth of the mycelium develops. After a few days the length of the grownout mycelium filaments was determined. Calculated were the EC_{50} value, the concentration of the DMI at which the length of the grown-out mycelium filaments is one-half of the outgrowth in the absence of DMI. The DMIs used were Fenarimol, Etaconazole and Imazalil. By way of check, the inhibition of growth by Benomyl was also looked at. This product inhibits mycelium growth via a different, 14 dm independent, mechanism. The results of these experiments are shown in Table VI and FIG. 6.

	N402	AB2-2	AB-D1	AB-D1.15
Estimated 14DM copy number	"1"	"1"	"1" + ≧10	"1" + ≧10
Estimated cprA copy number	1	≧10	1	≧10
Benomyl	2.27	3.05	1.93	3.22
	100%	134%	85%	142%
Fenarimol	1.73	3.88	11.50	58.58
	100%	224%	665%	3270%
Etaconazole	0.45	0.95	2.25	9.10
	100%	211%	500%	2022%
Imazalil	6.43	9.86	17.57	66.27
	100%	153%	273%	1030%

Table VI. Lanosterol 14α-demethylase inhibition (DMI) EC_{50} values expressed in ppm. Estimated 14 DM copy number: "1" wildtype *A. niger* 14 DM gene, =: extra copies of the *P. italicum* 14 DM gene.

Media and solutions

50×AspA (1 l.)

300 g $NaNO_3$
26 g KCl
76 g KOH_2PO_4
18 ml KOH (10M)

1000 * Spore elements (100 ml)

2.2 g $ZnSO_4.7H_2O$
1.1 g H_3BO_3
0.5 g $MnCl_2.4H_2O$
0.5 g $FeSO_4.7H_2O$
0.17 g $CoCl_2.6H_2O$
0.16 g $CuSO_4.5H_2O$
0.15 g $Na_2MoO_4.2H_2O$
5.0 g EDTA

Minimal medium (500 ml)

10 ml 50 * AspA
10 ml 50% glucose
0.5 ml 1000 * spore elements
1.0 ml 1M $MgSO_4$

Complete medium

10 ml 50 * AspA
10 ml 50% glucose
0.5 ml 1000 * spore elements
1.0 ml 1M $MgSO_4$
5.0 ml 10% Cas-amino acids
25 ml 10% Yeast extract

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 9

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3701 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
TGATAACTCC TCAGCAAATC GGAGTAAACA GAAGGACAAG TCATTGGAGT ACTAAGTAGC      60
TCCGTGTCAG AGACCCGGAC AGGATCAGCT TCTCCGAACC CGAGACTCCG GGCGAAAAGG     120
CCACCATCGC TCAGGCTACC ACCTGTGTTC CTTCCGTCGA TCGTCCTCCC TCGTTTCCGG     180
CTCACGGCCC CCCAAATTAT TGCGGTCTGC TTAGCAGTGG GTTCGGCCTC TCTGTTCTTC     240
CTGGATCACA CCACGGCTTA CTTTCTTATC CTTTTCCTTT TCCTTTCTTC CTTTCTTCCT     300
GTTCTCCTTT CTTCCTTTCC ACCCCCTTCT TTCTTTTAAC CCCATAGCGT CATTCTTTCT     360
TCCGTTTTAT CTTTGGTTTT GGGACGCCGC CACCTTATCT CGGTTCCTGC CTCGGTCTCC     420
```

-continued

```
GGTGATCGCA CCTGGATAGG CTAAGCGTAG GGAGGTGTGA CATTCTTCTT TCACCTCCTC   480
TCCTTTTCCC GCCTCACTCC GTTCAATCCC CCGCTCCACC CTTTCAGACT CGCCATCGTA   540
TCAAGTCGGG GCCTTTGCTT GCGCCGCTGA ACAGCCTCAC CATGGCGCAA CTCGATACCC   600
TCGATCTGGT GGTCCTGGCG GTGCTTTTGG TGGGTAGCGT GGCCTACTTC ACCAAGGGCA   660
CCTACTGGGC AGTTGCAAAG ACCCGTATGC CTCTACCGGC CCCGCGGATG AACGGCGCCG   720
CTAAGGCTGG CAAGACTCGG AACATCATTG AGAAGATGGA AGAAACGGGC AAGAATTGTG   780
TTATTTTCTA CGGATCGCAA ACTGGAACCG CTGAGGACTA CGCCTCCAGA TTGGCCAAGG   840
AAGGATCTCA GCGCTTCGGC CTCAAGACCA TGGTGGCTGA CCTCGAGGAA TACGACTATG   900
AGAACCTGGA CCAATTCCCG GAGGACAAGG TTGCGTTTTT CGTGCTCGCC ACCTACGGAG   960
AGGGTGAGCC TACGGATAAT GCTGTTGAGT TCTACCAGTT CTTCACCGGT GACGACGTTG  1020
CTTTTGAGAG CGCGTCCGCG GACGAGAAGC CTCTGTCCAA GCTGAAGTAT GTTGCTTTCG  1080
GTCTGGGTAA CAACACTTAT GAGCACTACA ACGCCATGGT TCGTCAAGTC GATGCTGCTT  1140
TCCAGAAGCT CGGGCCGCAG CGTATTGGTT CTGCTGGCGA GGGTGATGAC GGTGCCGGTA  1200
CAATGGAAGA AGACTTCTTG GCCTGGAAGG AGCCCATGTG GGCAGCACTG TCGGAGTCGA  1260
TGGATCTCGA AGAGCGTGAA GCGGTCTACG AACCTGTTTT CTGCGTCACC GAAAACGAGT  1320
CCCTGAGCCC TGAGGACGAG ACGGTCTATC TTGGAGAGCC CACCCAGAGC CACCTTCAGG  1380
GTACTCCCAA AGGCCCGTAC TCTGCGCACA ACCCCTTTAT CGCCCCTATT GCCGAATCTC  1440
GTGAGCTTTT CACCGTCAAG GATCGCAACT GTCTGCACAT GGAAATTAGC ATCGCTGGAA  1500
GTAACTTGTC CTACCAGACT GGTGACCACA TCGCTGTTTG GCCCACAAAC GCTGGTGCCG  1560
AAGTGGATCG GTTCCTTCAG GTCTTCGGTC TCGAGGGCAA GCGTGATTCG GTCATCAACA  1620
TCAAGGGTAT CGATGTTACG GCCAAGGTCC CAATCCCGAC CCCGACCACG TACGATGCCG  1680
CTGTTCGGTA CTATATGGAA GTCTGCGCCC CTGTGTCCCG TCAGTTTGTA GCCACTCTGG  1740
CCGCGTTCGC TCCGATGAGG AAAGCAAGGC AGAGATTGTG CGTCTGGGTA GCACAAGGAC  1800
TATTTCCACG AGAAGGTCAC CAACCAATGC TTCAACATGC CCAGGCTCTT CAGAGCATCA  1860
CGTCCAAGCC TTTCTCTGCT GTTCCGTTCT CTCTGCTTAT TGAAGGCATT ACGAAGCTGC  1920
AGCCTCGCTA CTACTCGATC TCTTCGTCCT CCCTTGTCCA GAAGGACAAG ATCAGCATCA  1980
CGGCCGTTGT GGAATCTGTT CGTCTGCCCG GTGCCTCTCA CATGGTGAAG GGTGTGACTA  2040
CGAATTATCT CCTCGCGCTC AAGCAGAAGC AGAACGGGCG ATCCCTCTCC CGACCCTCAC  2100
GGCTTGACTT ACTCCATCAC GGTCCCCGGA ACAAGTACGA CGGTATCCAC GTTCCCGTGC  2160
ATGTTCGCCA CTCGAACTTC AAGCTGCCCT CTGATCCCTC TCGGCCCATT ATCATGGTTG  2220
GTCCTGGTAC TGGTGTTGCT CCTTTCCGTG GTTTCATTCA GGAACGTGCT GCTTTGGCGG  2280
CCAAGGGCGA GAAGGTTGGA CCCACTGTTC TCTTCTTCGG TTGCCGCAAG AGTGACGAGG  2340
ATTTCTTGTA CAAGGATGAA TGGAAGGTAA GATATCTTTT TTTCTTTTCC GCAGCTACCT  2400
TCATACATCT CGGATGCTAA CATATCGCGA TTCGCAGACC TATCAGGACC AGCTTGGAGA  2460
CAACTTGAAG ATCATCACTG CGTTCTCGCG TGAGGGTCCT CAGAAGGTCT ACGTTCAGCA  2520
CAGACTCCGC GAGCACTCCG AACTTGTCAG CGACCTTCTG AAGCAGAAAG CTACCTTCTA  2580
CGTCTGTGGT GACGCTGCAA ACATGGCTCG CGAGGTTAAC CTTGTGCTTG GCCAGATCAT  2640
TGCTGCGCAG CGTGGTCTGC CCGCCGAGAA GGGCGAAGAA ATGGTCAAGC ACATGCGTAG  2700
ACGTGGACGC TACCAGGAAG ATGTGTGGTC ATAATCTTTC AATGCATCGA CTTTTCTTTC  2760
TTGTCTATCA CGACGGCCTT CTCGATCCAT TATTTTATTT AACGCCTAGA TGATCTTTGC  2820
```

-continued

```
ATATATACTC CGCTGATTTT GCCTATTCAT CTGTTTTGCT TGGCGTGGTT TATGTATGCC   2880
TAGTTTATTT GTTTTGTGCA CCGACCGGCC AGCCACACAT TGAAGTGGCT TGAGCATGAG   2940
TGCGGTAGCC AGTGTCGAAA GAACAGGATA GACGATCATG ATTATTGCGG GAACATGTTA   3000
TGCCATTCTG GGCATATTGA TATCTGGTTG CATGAGCCCA GAGGATACGA AAAGATGAAT   3060
CCATATTTAA TTTGCACAAT ACTTTTCGCC TTCTTCATCT AGTAATTAAA TTAATTGAGC   3120
ACTGACCGAA CGAGCTGACA CCTGCTGCTC GGAATAGCCG ACAACGCATT GACGTGCAAG   3180
AGATGCATAA TCATTACAAT CAACAAGTAG ACTGGTAACT AAATCACTGA ATACTACAGT   3240
TACTGCCTAC TTTCAGCCAA AAAGTAATAC TGAAGATTTC GGGGAATCAA ATAGAAGAAA   3300
CATGCATAAG CCCAACCTCG GCAATACCGG GAGTTAAGCA CAGTAACCAA AACCAAACCA   3360
AACTAGAACC GGCGCGCGAC CAGTGACCCA TCGTCATTCC CGGTATCAGC AGTTCAGTCA   3420
GACTGGCTGG CTAGCCCGAA CCCAACTGCC GCAATCATCC ATCCATCCTC AACCCGCCCC   3480
TCCCATGCCA ACCTCTCTAC TCCGCAGAGC GAGGGACAAA AAAATGAGAT GCAGCAATTA   3540
ACCACGATAA TCTAGCAAAA AGAAAGTTAG AAGCCGGAAG AACATACATA TCGCTTTTAC   3600
CGCTGTTCGA CTGCGACGAC GGGTCTTGAG AGCAGTTCCG CCACGTGGGC GAAAAGCTGG   3660
ACTGCACACT ACTTACGCTA CCCTACGCTA CCTCGGTACC C                       3701
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 693 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ala Gln Leu Asp Thr Leu Asp Leu Val Val Leu Ala Val Leu Leu
1               5                  10                  15

Val Gly Ser Val Ala Tyr Phe Thr Lys Gly Thr Tyr Trp Ala Val Ala
            20                  25                  30

Lys Thr Arg Met Pro Leu Pro Ala Pro Arg Met Asn Gly Ala Ala Lys
        35                  40                  45

Ala Gly Lys Thr Arg Asn Ile Ile Glu Lys Met Glu Glu Thr Gly Lys
    50                  55                  60

Asn Cys Val Ile Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Asp Tyr
65                  70                  75                  80

Ala Ser Arg Leu Ala Lys Glu Gly Ser Gln Arg Phe Gly Leu Lys Thr
                85                  90                  95

Met Val Ala Asp Leu Glu Glu Tyr Asp Tyr Glu Asn Leu Asp Gln Phe
            100                 105                 110

Pro Glu Asp Lys Val Ala Phe Phe Val Leu Ala Thr Tyr Gly Glu Gly
        115                 120                 125

Glu Pro Thr Asp Asn Ala Val Glu Phe Tyr Gln Phe Phe Thr Gly Asp
    130                 135                 140

Asp Val Ala Phe Glu Ser Ala Ser Ala Asp Glu Lys Pro Leu Ser Lys
145                 150                 155                 160

Leu Lys Tyr Val Ala Phe Gly Leu Gly Asn Asn Thr Tyr Glu His Tyr
                165                 170                 175
```

-continued

Asn Ala Met Val Arg Gln Val Asp Ala Ala Phe Gln Lys Leu Gly Pro
180 185 190

Gln Arg Ile Gly Ser Ala Gly Glu Gly Asp Asp Gly Ala Gly Thr Met
195 200 205

Glu Glu Asp Phe Leu Ala Trp Lys Glu Pro Met Trp Ala Ala Leu Ser
210 215 220

Glu Ser Met Asp Leu Glu Glu Arg Gln Ala Val Tyr Glu Pro Val Phe
225 230 235 240

Cys Val Thr Glu Asn Glu Ser Leu Ser Pro Glu Asp Glu Thr Val Tyr
245 250 255

Leu Gly Glu Pro Thr Gln Ser His Leu Gln Gly Thr Pro Lys Gly Pro
260 265 270

Tyr Ser Ala His Asn Pro Phe Ile Ala Pro Ile Ala Glu Ser Arg Glu
275 280 285

Leu Phe Thr Val Lys Asp Arg Asn Cys Leu His Met Glu Ile Ser Ile
290 295 300

Ala Gly Ser Asn Leu Ser Tyr Gln Thr Gly Asp His Ile Ala Val Trp
305 310 315 320

Pro Thr Asn Ala Gly Ala Glu Val Asp Arg Phe Leu Gln Val Phe Gly
325 330 335

Leu Glu Gly Lys Arg Asp Ser Val Ile Asn Ile Lys Gly Ile Asp Val
340 345 350

Thr Ala Lys Val Pro Ile Pro Thr Pro Thr Thr Tyr Asp Ala Ala Val
355 360 365

Arg Tyr Tyr Met Glu Val Cys Ala Pro Val Ser Arg Gln Phe Val Ala
370 375 380

Thr Leu Ala Ala Phe Ala Pro Met Arg Lys Ala Arg Gln Arg Leu Cys
385 390 395 400

Val Trp Val Ala Gln Gly Leu Phe Pro Arg Glu Gly His Gln Pro Met
405 410 415

Leu Gln His Ala Gln Ala Leu Gln Ser Ile Thr Ser Lys Pro Phe Ser
420 425 430

Ala Val Pro Phe Ser Leu Leu Ile Glu Gly Ile Thr Lys Leu Gln Pro
435 440 445

Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Val Gln Lys Asp Lys Ile
450 455 460

Ser Ile Thr Ala Val Val Glu Ser Val Arg Leu Pro Gly Ala Ser His
465 470 475 480

Met Val Lys Gly Val Thr Thr Asn Tyr Leu Leu Ala Leu Lys Gln Lys
485 490 495

Gln Asn Gly Arg Ser Leu Ser Arg Pro Ser Arg Leu Asp Leu Leu His
500 505 510

His Gly Pro Arg Asn Lys Tyr Asp Gly Ile His Val Pro Val His Val
515 520 525

Arg His Ser Asn Phe Lys Leu Pro Ser Asp Pro Ser Arg Pro Ile Ile
530 535 540

Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Arg Gly Phe Ile Gln
545 550 555 560

Glu Arg Ala Ala Leu Ala Ala Lys Gly Glu Lys Val Gly Pro Thr Val
565 570 575

Leu Phe Phe Gly Cys Arg Lys Ser Asp Glu Asp Phe Leu Tyr Lys Asp
580 585 590

Glu Trp Lys Thr Tyr Gln Asp Gln Leu Gly Asp Asn Leu Lys Ile Ile
595 600 605

-continued

Thr Ala Phe Ser Arg Glu Gly Pro Gln Lys Val Tyr Val Gln His Arg
610 615 620

Leu Arg Glu His Ser Glu Leu Val Ser Asp Leu Leu Lys Gln Lys Ala
625 630 635 640

Thr Phe Tyr Val Cys Gly Asp Ala Ala Asn Met Ala Arg Glu Val Asn
645 650 655

Leu Val Leu Gly Gln Ile Ile Ala Ala Gln Arg Gly Leu Pro Ala Glu
660 665 670

Lys Gly Glu Glu Met Val Lys His Met Arg Arg Arg Gly Arg Tyr Gln
675 680 685

Glu Asp Val Trp Ser
690

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CCGGAATTCC ARACNGGNAC NGCNGARGA 29

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCGGAATTCG GNGANCCNAC NGAYAAYGC 29

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CGCGGATCCG GNCCNAYNAD DATNAC 26

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

-continued

(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(i i) MOLECULE TYPE: other nucleic acid

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGGGATCCT SYTGNACRTA NACYTT 26

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(i i) MOLECULE TYPE: other nucleic acid

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGCGGATCCG GNCCDATCAT DATNAC 26

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(i i) MOLECULE TYPE: other nucleic acid

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CCACGCTACC CAC 13

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(i i) MOLECULE TYPE: other nucleic acid

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAGGCATACG GGTC 14

We claim:

1. A recombinant DNA molecule comprising one of

(a) a nucleic acid sequence encoding a cytochrome P450 oxidoreductase having an amino acid sequence as shown in SEQ ID NO:2,

(b) a nucleic acid sequence which hybridizes under stringency conditions of 56° C. and 6X SSC with the nucleic acid sequence of (a);

(c) a nucleic acid sequence complementary with the nucleic acid sequence of (a), and

(d) a nucleic acid sequence complementary with the nucleic acid sequence of (b).

2. An RNA molecule coding for a cytochrome P450 oxidoreductase, which is derived from a DNA molecule according to claim 1.

3. A polypeptide having cytochrome P450 oxidoreductase activity, said polypeptide being coded for by a recombinant DNA molecule according to claim 1.

4. A transformed host cell which is at least transformed with a DNA molecule according to claim 1 and which is derived from a filamentous fungus.

5. A host cell according to claim 4 which is derived from a filamentous ascomycete.

6. A host cell according to claim 5 which is derived from the genus Aspergillus.

7. A host cell according to claim 6 which is derived from *Aspergillus Niger*.

8. A host cell according to claim 4 which is further transformed with a DNA molecule which codes for a p 450 cytochrome protein.

9. A host cell according to claim 8, wherein the P450 cytochrome protein is coded for by one of

(a) a DNA molecule as shown in SEQ ID NO:1, and

(b) a DNA molecule which hybridizes under moderately stringent conditions with a DNA molecule according to SEQ ID NO:1.

10. A process for enzymatic conversion of a substrate, comprising contacting the substrate with a filamentous fungus capable of converting said substrate and optionally recovering the conversion product formed, wherein said filamentous fungus is a recombinant filamentous fungus transformed with and expressing a recombinant DNA molecule according to claim 1.

11. A process for enzymatic conversion of a substrate, comprising contacting the substrate with an enzyme capable of converting said substrate and optionally recovering the conversion product formed, wherein said enzyme is a cytochrome P450 oxidoreductase according to claim 3.

12. A recombinant DNA molecule comprising one of

(a) a part of nucleic acid sequence SEQ ID NO:1 that encodes a cytochrome P450 oxidoreductase,

(b) a nucleic acid sequence which hybridizes under stringency conditions of 56° C. and 6X SSC with the nucleic acid sequence of (a),

(c) a nucleic acid sequence complementary with the nucleic acid sequence of (a), and

(d) a nucleic acid sequence complementary with the nucleic acid sequence of (b).

13. A process for enzymatic conversion of a substrate, comprising contacting the substrate with a filamentous fungus capable of converting said substrate and optionally recovering the conversion product formed, wherein said filamentous fungus is a recombinant filamentous fungus transformed with and expressing a recombinant DNA molecule according to claim 12.

* * * * *